Akesson, Jesper; Gathergood, John; Quispe-Torreblanca, Edika

Working Paper

Preventing payments fraud in the FinTech era: New evidence from a behavioural experiment

CeDEx Discussion Paper Series, No. 2023-08

Provided in Cooperation with:
The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx)

Suggested Citation: Akesson, Jesper; Gathergood, John; Quispe-Torreblanca, Edika (2023) : Preventing payments fraud in the FinTech era: New evidence from a behavioural experiment, CeDEx Discussion Paper Series, No. 2023-08, The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx), Nottingham

This Version is available at:
https://hdl.handle.net/10419/284283

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

University of Nottingham
UK | CHINA | MALAYSIA

Discussion Paper No. 2023-08

Jesper Akesson,
John Gathergood and
Edika Quispe-Torreblanca

Preventing Payments Fraud in the FinTech Era: New Evidence from a Behavioural Experiment

August 2023

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Samantha Stapleford-Allen
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 74 86214
Samantha.Stapleford-Allen@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

Preventing Payments Fraud in the FinTech Era: New Evidence from a Behavioural Experiment

Jesper Akesson, John Gathergood & Edika Quispe-Torreblanca*

August 5, 2023

Abstract

Innovation in financial technology has granted consumers increased access to faster, more convenient payment services. This development has, however, also given rise to Authorised Push Payment (APP) fraud, where consumers are unwittingly manipulated into authorising transactions to counterfeit parties, such as fake online sellers. The annual costs of APP fraud are growing, and for example total more than £0.5bn in the United Kingdom alone. In this paper, we present the results from an online experiment that tests interventions designed to reduce the likelihood that consumers fall for APP fraud. These interventions were presented to consumers within a mobile bank application, and for instance, involved presenting warnings and increasing the salience of calls-to-action. Our analysis shows that redesigned calls-to-action can dramatically reduce fraud success rates, whereas traditional behavioural and risk-based warnings have much weaker effects. Our results show how redesigning consumer journeys can potentially reduce fraud prevalence.

KEYWORDS: fraud, financial technology, behavioural science
JEL CODES: G21, G41, G5

*Jesper Akesson is with The Behaviouralist, John Gathergood is with the University of Nottingham, Edika Quispe-Torreblanca is with the University of Leeds. The research presented in this paper was undertaken in collaboration with the Open Banking Implementation Entity in the United Kingdom. The Behaviouralist prepared a report for the Open Banking Implementation Entity "Using Behavioural Insights and Experimentation to Prevent APP Fraud", which presents the experimental results reported in this paper. Any views or opinions expressed in this paper do not necessarily reflect those of the Open Banking Implementation Entity. All mistakes are the authors' own.

1 Introduction

The rapid development of financial technology over the past few years has revolutionized the payments industry. Consumers now have access to a wide range of payment options that are cheaper, faster, and more convenient than traditional banking methods. Mobile payment apps, for instance, enable consumers to push payments to counterparty bank accounts directly, allowing sellers to avoid the cost of interchange fees, among other benefits (Hayashi, 2012; He et al., 2023). Open banking regulations have facilitated the rise of new forms of payments in the United Kingdom, followed by Australia and the United States (Ziegler, 2021). These new forms of payments are also seen as a way to increase competition in the payments industry, which is traditionally dominated by banks. However, there is concern that some innovations in financial technology might lead to, or facilitate, fraud (Button and Cross, 2017; Griffin et al., 2023).

Despite their many benefits, these new payment methods open up new opportunities for fraudsters to deceive consumers. One form of such deception is the Authorised Push Payment (APP) scam, which entails fraudsters posing as online merchants, or banks, or even as relatives in need of urgent financial assistance, and persuading consumers to transfer money to their personal accounts for non-existent or non-delivered products, failed account transfers, or fabricated emergencies. In these instances, unwitting consumers inadvertently transfer funds, which later turn out to be the result of fraud.[1]

This type of fraud can be devastating to victims, leading to significant financial losses and damage to their credit scores. According to the UK Payment Systems Regulator, in 2023, over 120,000 UK consumers fell foul of this form of fraud, collectively losing approximately £0.5bn (PSR, 2023). The prevalence of APP fraud underscores the urgency of implementing interventions aimed at reducing its incidence. As such, the drive to mitigate the occur-

[1] APP scams are typically classified into 1) malicious payee (purchase scam, investment scam, romance scam, and advance-fee scam) and 2) malicious redirection (invoice & mandate scam, CEO fraud, and impersonation).

rence of APP fraud is a major concern for regulators, financial institutions, and consumer organisations that hold a vested interest in the payments sector.

In response to this challenge, this paper uses an online experiment to test a range of behavioural interventions that aim to reduce the likelihood of consumers falling for APP fraud. We designed an experiment in which subjects faced a number of payment tasks (e.g., paying for the purchase of an online good), with some tasks proving to be legitimate while others fraudulent. Participants were incentivised to make legitimate payments and avoid falling for fraudulent payments. In the experiment, fraudulent payments contained information cues from which consumers could deduce that they were not legitimate. Our experiment incorporates a user interface based upon existing payment journeys adopted by financial services providers in the UK as the basis for the design of the control journey. We introduce eight different payment journey treatments designed to reduce the likelihood of subjects making fraudulent payments.

Our main finding is that marketing-inspired "call to action" (CTA) interventions were most effective at reducing rates of fraudulent payments, in comparison to interventions designed to appeal to loss aversion of social norms. We show that the most effective interventions were those that incorporated salient calls to action that provided users with options to cancel or postpone payments and combined warnings only for high-risk payments. These interventions had dramatic effects on the share that fell for fraud. Participants in the 'Risk-based + CTA' treatment were 81% less likely to fall for fraud than those in the control group. In comparison, the use of behavioural warnings reduced the share of participants that fell for fraud by 18%. However, the effectiveness of behavioural warnings decayed with increasing exposure to fraud, as the differences in the share of participants who fell for fraud were no longer significant when fraud occurred in the third payment scenario compared to the first, suggesting that participants became desensitized to these warnings. In contrast, CTA interventions were able to sustain their significant effects and even enhance them when

participants encountered fraud in the third scenario, resulting in a 94% decrease in fraudulent payments.

A feature of our study is the use of an online experiment to test the effectiveness of interventions to reduce fraud. The feasibility of real-world testing of interventions is compromised by the low rates of fraud as a share of total payments, and incomplete detection of fraud.[2] An additional feature of our study is that our experiment, which was administered via an online survey, tested a broad set of interventions.

We recruited a nationally representative sample of approximately 10,000 UK adults (with internet access) to take part in the study. Participants were randomly allocated to eight main experimental conditions. Individuals allocated to the different conditions were shown slightly different versions of a bank app. Those in the first group (the control condition) were shown warning interventions that closely resemble approaches currently observed in the market. Those in the second group (the behavioural condition) were shown the same app, with the addition of warnings that appealed to participants' loss aversion and social norms. Those in the third group (the Call to Action, or CTA, condition) were also shown an app closely resembling the 'control' app, but with the addition of salient calls to action (or buttons) that offered users the opportunity to cancel or postpone payments. Those in the fourth group (the behavioural and CTA condition) were shown an app that combined the features of the behavioural and CTA apps. Further, those in groups five to eight were shown apps with the same additional features (i.e., adding behavioural warnings or CTAs), but with the key difference that the apps took a risk-based approach. This meant that warnings were primarily triggered when payments were deemed to be suspicious.

Our study was undertaken in collaboration with the Open Banking Implementation Entity

[2] A real-world experiment in a live payments app might require a sample of many millions of payments in order to be statistically powered, and would rely on accurate detection and reporting of the outcome variable (cases of fraudulent payments). In the online experimental setting, it is feasible to imitate the online screen of a consumer payments journey, and create incentive-compatible payment structures in the experiment, thereby offering a greater degree of external validity.

(OBIE), which commissioned and funded the study. OBIE was established in 2016 by the UK's competition regulator (the Competition and Markets Authority) to established the technology standards for open banking in the UK. OBIE is funded by a levy on the nine largest personal current account providers in the UK, with its budget and workplans determined by the CMA.[3] This study formed part of a broader programme of research focused on improving consumer experience of payments journeys.

Our study contributes to the literature on consumer fraud and deception. Previous studies have shown that fraudsters tend to impersonate those in authority (as in Luo et al., 2013), often using online or telephone communication to cloak their appearance (Chang and Chong, 2010), imitating the pro-forma and templates of legitimate communications and gain access to the personal details of victim's to appear as an informed party (Finn and Jakobsson, 2007). Fraudsters also tend to deliver an urgent or dramatic key messages such as an unbelievable bargain or an urgent family need, in order to motivate the victim to act, as in the elaboration likelihood model of Petty et al. (1986). Susceptibility to fraud is likely to be higher among those who lack self-control (Wilsem, 2013; Dickman, 1990) and exhibit impulsive personalities (Pattinson et al., 2011; Chen et al., 2017; Reisig and Holtfreter (2013)). Recent research considers the potential for introducing regulatory standards to mitigate APP fraud (Taylor and Galica, 2020; Maher, 2021; McIlroy and Sethi-Smith, 2021; Dahlgreen, 2021), including the use of new technologies to aid its prevention (Ma et al., 2023). Yet, despite the growing threat of APP fraud, there is currently limited evidence on which interventions are most effective in preventing it. This study addresses this gap by testing a range of behavioural interventions in an online experiment with a nationally representative sample of UK adults.

More broadly, our paper contributes to the emerging literature on the costs and benefits of innovation in financial technology. Recent research examines how fintech has affected payments services and lending practices (Allen et al., 2021; Berg et al., 2022; Di Maggio et al.,

[3]These institutions are AIBG, Bank of Ireland, Barclays, Danske, HSBC, Lloyds Banking Group, Nationwide, RBS and Santander.

2022; Yang and Zhang, 2022), including the potential benefits for underserved consumers and for financial stability (Buchak et al., 2021; Johnson et al., 2023; Parlour et al., 2022; Philippon, 2016).

Our findings inform the ongoing evolution of open banking regulation in the UK. The growing losses caused by APP fraud prompted regulators, as well as the financial industry, to take a number of important steps to reduce fraud, such as improving payee identity verification.[4] As more consumers turn to online transactions, the risk of fraud increases, and regulatory frameworks have been established to mitigate these risks. However, there is limited empirical evidence on the effectiveness of different fraud prevention measures, and our study addresses this gap. Furthermore, our findings can inform the ongoing evolution of open banking regulation in the UK, building on the steps already taken to combat APP fraud, such as the establishment of the CRM code and the introduction of CoP services. Overall, our study contributes to the literature on consumer fraud and deception, and its findings have important implications for the payments industry.

The paper is structured as follows: Section 2 describes the experimental design and the interventions tested. Section 3 describes our empirical strategy. Section 4 presents the results of the experiment. Section 5 discusses the implications of the findings and concludes the paper.

[4]In February 2018, the Payment Systems Regulator established a steering group to lead the development of an industry code—known as the Contingent Reimbursement Model (CRM)—that deals with APP fraud. The CRM code (which is now in effect) sets out consumer protection standards related to the reimbursement of victims and aims to reduce the prevalence of APP scams. Another important step that has been taken, at least in part, to reduce the prevalence of APP fraud is the introduction of Confirmation of Payee (CoP) services. These services help customers verify whether the account name they entered matches the registered account name. This innovation helps to avoid some forms of fraud in which the fraudsters acts as an imposter while providing their own bank details.

2 Experimental Design

In this section, we outline the selection process for the payment scenarios, detail the design of the customer journeys implemented to prevent APP fraud, and describe the methodology of the online experiment, including our sample of participants.

2.1 Payment Scenarios

The objective of this study is to investigate APP fraud, where consumers are deceived into authorising a transfer of money to a fraudster's account. Such fraud can be perpetrated via phone, online, or in-person communication, with varying mechanisms and types of scams, causing substantial financial losses for victims.

To recreate online payment scenarios that resemble real-life situations, we began by identifying the most prevalent types of APP scams. In Table A1, we show the most common types of APP scams, along with the frequency and value of losses.

We narrowed down the initial list by applying two key criteria. We prioritised scams that accounted for the highest volume of losses and could be recreated in an online environment within a reasonable timeframe (without relying solely on phone calls or in-person communication). After applying these criteria, we were left with a selection of four types of scams deemed most suitable for testing: invoice and mandate scams, impersonation scams involving bank staff or police, purchase scams, and investment scams.

After narrowing down the selection, we researched the characteristics and tell-tale signs of each scam type and developed a set of preliminary scenarios with the help of OBIE's fraud specialists.[5] These scenarios were refined based on feedback from stakeholders and tested in two pilot studies in which we presented participants with either the fraudulent or legitimate version of each scenario and asked them to rate the likelihood of each representing a fraudulent payment request.

[5]Table A2 shows the list of tell-tale signs of the final scam types we recreated.

Three payment scenarios were ultimately chosen for the experiments: an invoice scam for kitchen remodelling, a purchase scam (buying a laptop) on Facebook Marketplace, and an impersonation scam involving overdue self-declared taxes to HMRC. For each scenario, we developed fraudulent and legitimate versions with tell-tale signs of the corresponding scam type and elements of reassurance, respectively. Figure 1 presents an example of the evidence attached to the payment scenario involving the purchase of a laptop from a seller on the Facebook Marketplace. Figures A1 and A2 in Appendix B illustrate the evidence attached to the remaining scenarios.

2.2 Developing Customer Journeys

In collaboration with OBIE, we developed eight distinct payment journeys within bank apps for our experiment. These journeys were intended to reduce instances of APP fraud while also minimising customer friction. We also aimed to measure any potential secondary effects on customer experience and the proportion of legitimate payments that customers completed. To inform the design of these journeys, we conducted research on current bank practices for preventing APP fraud, the characteristics of such fraud, and the existing literature on behavioural economics.

Our experiment evaluated the impact of the following interventions on the payment journeys:

1. Risk-based approach: This involves gathering additional information about the payment and displaying warnings only for high-risk payments. By targeting payments based on their level of risk, a risk-based approach could minimise disruption to legitimate payments while reducing fraud.

2. Amended Calls to Action (CTAs): In this approach, there are more buttons within the app for participants to cancel, save payments for later, or call the bank. Offering a clear exit can reduce fraud when customers cannot stop the payment process but distrust fraudsters, and delaying transactions can prevent fraud by limiting the decision-making

time that fraudsters exploit with time-limited messages.

3. Behavioural interventions: These incorporate text messages that leverage loss aversion by highlighting the potential loss of a given amount of money if the user proceeds with the payment.

We also found evidence suggesting the importance of clear, personalised, and differentiated risk communications. Thus, we developed two branches for our experiment. In the first branch, also referred to as the control branch, participants were allocated to a typical bank transfer journey currently available on the market. In the second branch, known as the risk-based branch, we incorporated a risk-assessment algorithm that determined when warnings should be triggered. We tested three interventions within each branch, which are described below.

- Control group
- Control group + behavioural interventions (Figure 3)
- Control group + CTAs (Figure 4)
- Control group + behavioural interventions + CTAs (Figure 5)
- Risk-based group (Figure 6)
- Risk-based group + behavioural interventions (Figure 7)
- Risk-based group + CTAs (Figure 8)
- Risk-based groups + behavioural interventions + CTAs (Figure 9)

Thus, participants in our study could be assigned to one of eight different customer journeys. Furthermore, the risk-based group was split into two subgroups: high accuracy, which did not misclassify high-risk scenarios as low-risk or vice versa, and low accuracy, which produced some false positives by classifying legitimate scenarios as risky. A visualisation of the experimental design can be observed in Figure 2.

2.3 Experiment

The study was carried out on the Qualtrics survey platform, using a nationally representative sample of 15,888 adult UK participants who were recruited through Panelbase and Prolific Academic. The Panelbase sample was representative in terms of age, gender, and location, while the Prolific sample was representative in terms of age, gender, and ethnicity. Prior to beginning the study, participants were asked to provide information on their age, gender, location, household income, online banking usage, and banks they use.

After providing this information, participants were presented with the instructions to complete three payment tasks, as shown in Figure B1. They were then asked to respond to comprehension questions related to the study, with nearly 90% of participants correctly answering at least four of these questions.

The participants were then shown one of the three payment scenarios outlined in Section 2.1, with one payment request being fraudulent and the other two being legitimate but unbeknown to the participants.[6] The presentation order of scenarios, participants' assignment to one of the eight banking journeys, as well as the display of either a fraudulent or legitimate scenario version to a participant, were all randomized. Participants were incentivized to make legitimate payments and avoid fraudulent requests by using a mobile banking app that was semi-interactive, and they were asked to complete a survey regarding their experience with the app after they had completed all three payment scenarios.[7]

The primary outcomes of the study were whether participants made fraudulent or legitimate payments, while secondary outcomes included participant perceptions of the app, their willingness to recommend the app, and the time it took to complete a payment.

Table C1 in the Online Appendix provides descriptive statistics on the 8958 participants who

[6]The description and/or the evidence provided were modified to represent a fraudulent scenario that shared most of the same information as the legitimate scenario.

[7]The banking app that participants were shown was inspired by a recent version of an existing banking app.

completed all three scenarios, while Tables C2 through C4 in the Online Appendix show the statistical balance in all three randomizations performed.

3 Empirical Strategy

We estimate the likelihood of fraudulent and legitimate payments using linear probability models. Our outcomes of interest, described in Equations 1 and 2, are represented by the dummy variable $Fraudulent_i$, which takes the value of 1 if the participant made a fraudulent payment and 0 otherwise, and the semi-continuous variable $Legitimate_i$, which measures the share of legitimate payments made and can take the values 0, 0.5, or 1, as participants face two legitimate payment scenarios. We analyse the experimental results at the individual level i, with the eight customer journeys in our experiment as the key independent variables represented by seven dummy variables. The omitted category $Control_i$ represents a typical bank payment journey without any intervention.

$$\begin{aligned} Fraudulent_i = b_0 + b_1 ControlBeh_i + b_2 ControlCTA_i + b_3 ControlBehCTA_i + \\ b_4 Riski + b_5 RiskBeh_i + b_6 RiskCTA_i + b_7 RiskBehCTA_i + \epsilon_i \end{aligned} \tag{1}$$

$$\begin{aligned} Legitimate_i = b_0 + b_1 ControlBeh_i + b_2 ControlCTA_i + b_3 ControlBehCTA_i + \\ b_4 Risk_i + b_5 RiskBeh_i + b_6 RiskCTA_i + b_7 RiskBehCTA_i + \epsilon_i \end{aligned} \tag{2}$$

The results are organised into three primary sections. The first section presents the treatment effects of each intervention and discusses potential interactions among the interventions. We conduct robustness and sensitivity tests in the second and third sections, respectively. Specifically, we report the treatment effects by payment scenario, as well as the effects after accounting for demographic factors and correcting for multiple hypothesis testing. As part of sensitivity analysis, we investigate heterogeneity across demographic groups and examine the effect of the order of the fraud scenario. Finally, we report the effects on customer

satisfaction and time spent across all payment scenarios.

4 Results

4.1 Main Results

Table 1 and Figure 10 present the main findings of our study, examining the effects of various payment journeys on fraudulent and legitimate payments. First, we address the question of which of the eight experimental groups had the greatest effect on fraudulent payments. We observe that in the absence of any intervention, the control group had an average of 22% of participants making fraudulent payments and 57% completing legitimate payments, as indicated by the regression constants. Among the seven journeys, five exhibited statistically significant effects on the proportion of participants making fraudulent payments compared to the control group. The largest effect was observed in the 'Risk-based + CTA' group, which had an 18-percentage-point decrease relative to the control group. The 'Risk-based + behavioural + CTA' group was the second most successful journey (with a 16-percentage-point effect), followed by the 'Control + behavioural + CTA' group (with a 14 percentage-points effect), and the 'Control + CTA' group (with a 12-percentage-point effect). These outcomes suggest that the CTA element had a significant impact on reducing fraudulent payments, regardless of whether it was paired with a risk-based or behavioural approach.

By introducing friction in the payment process, interventions aimed at reducing fraudulent payments may come at a cost of decreasing legitimate transactions. Customers may find it more difficult or time-consuming to complete their transactions, and they may also have a sense of distrust or suspicion as a result of the interventions, causing them to abandon their transactions altogether. We now move onto analysing this trade-off between combating fraudulent payments and preserving legitimate transactions. As shown in Column 2 of Table 1, the CTA payment journeys unintentionally reduced the proportion of non-fraudulent payments completed. In contrast, while the 'Risk-based' journey did not significantly reduce

fraudulent payments, it increased the share of legitimate payments completed by 8 percentage points. When combined with the behavioural intervention, the 'Risk-based + behavioural' journey resulted in a further 2-point increase in legitimate payments. The inadequacy of the risk-based approach to effectively mitigate fraud may suggest that the mere dissemination of information regarding the prevalence and risks associated with fraud may not suffice. Rather, individuals may need strategically placed prompts that remind them of their options and offer a convenient way to cancel payments when suspicious. It is worth noting that the risk-based approach's positive effect on legitimate payments could be attributed to the removal of redundant warnings. Thus, a more targeted approach to warning dissemination that is triggered during 'risky' transactions may prove more effective in increasing the proportion of legitimate payments.

To obtain a more precise estimate of the effect of each intervention and any potential interactions, we re-estimated the baseline models using dummy variables for each of the three interventions, as shown in Table C5. The results indicate that CTAs had a significant main effect, with a 12-percentage-point reduction in the likelihood of making fraudulent payments compared to the control group. Moreover, there was a significant interaction between CTAs and the risk-based approach, indicating that the risk-based approach further reduced the likelihood of making fraudulent payments by 7 percentage points when combined with CTAs. Thus, combining strategically positioned CTAs with targeted warning distribution may prove more efficient in mitigating fraud.

In consistency with the findings observed earlier in Table 1, the CTA interventions also had an unintended consequence of decreasing the proportion of legitimate payments completed. Specifically, the CTAs approach reduced the likelihood of legitimate payments by 14 percentage points with respect to the control group. In contrast, the risk-based interventions did not reduce fraudulent payments, but they did increase the share of legitimate payments completed by 8 percentage points.

The estimates outlined above demonstrate different trade-offs between reducing fraudulent payments and increasing the share of legitimate payments across interventions. Both the 'Risk-based' and 'Risk-based + behavioural' groups showed no unintended consequences, and instead, increased the proportion of legitimate payments. However, they did not significantly reduce fraudulent payments. In contrast, the CTA interventions were highly effective in reducing fraudulent payments but also reduced the share of legitimate payments completed. Despite the unintended consequences of the CTA interventions, the magnitude of the reduction in fraudulent payments was much larger than the unintended effect on legitimate payments. For instance, the 'Risk-based + CTA' journey reduced the percentage of fraudulent payments from 22% in the control group to 4%, while legitimate payments only decreased from 57% to 50%.

It is important to note that by informing participants at the beginning of the experiment that some payments may not be legitimate, they may have become more suspicious and cautious when completing their payment journeys. Consequently, we may be overestimating the negative impact of the treatments on the proportion of participants who completed legitimate journeys. Therefore, our estimates could be regarded as upper bounds of the impact of the interventions on legitimate payments.

Taken together, our findings suggest that well-placed CTA interventions have the most substantial impact on reducing fraudulent payments, whereas warnings regarding the prevalence of fraud were less effective when presented in isolation. Furthermore, a blend of risk-based and CTA interventions proved to be the most efficient method of mitigating fraud while limiting unintended consequences.

4.2 Robustness Tests

4.2.1 Intervention Effects on Payment Scenarios

The effectiveness of interventions may vary depending on the specific details of each payment scenario and the psychological factors involved in each scam. We conducted our first robustness test by evaluating whether the effects of our interventions were only apparent in particular payment scenarios.

We present the results in Table 2 for the likelihood of fraudulent payment and in Table 3 for the likelihood of legitimate payments. Each column in Table 2 represents a payment scenario, namely the fraudulent version of the invoice for the kitchen remodelling scenario, the laptop purchase scenario, and the HMRC tax payment scenario. Among the three payment scenarios, the fraudulent laptop scenario had the highest likelihood of deceiving participants, with a likelihood of 33%, which was twice as high as that of the fraudulent HMRC scenario (17%) and the fraudulent kitchen scenario (16%).

We also observe differences in treatment effectiveness across the payment scenarios. First, the behavioural intervention alone had no effect on the fraudulent HMRC scenario, indicating that the efficacy of this intervention is contingent on the underlying psychological factors involved in each particular scam. In the case of the purchase scam on Facebook Marketplace, for example, the use of loss aversion may have been more effective due to participants' emotional attachment to the idea of purchasing a new laptop. However, for the impersonation scam involving overdue taxes to HMRC, participants might have been more motivated by fear of legal consequences than by the potential loss of money, making this intervention less effective.

Second, the 'Risk + CTA' and 'Risk + behavioural + CTA' journeys had the largest effects overall across all payment scenarios, with the largest impact on reducing fraudulent payments in the laptop purchase scenario, with an effect size of 28 percentage points, doubling that of the other two scenarios. Considering that the laptop scenario had the highest percentage

of people making fraudulent payments (33%) in its respective control group, this reduction brings fraudulent payments to just 5%. Despite these variations in absolute effect size across payment scenarios, our findings indicate that the 'Risk + CTA' journey had the highest relative impact on reducing fraudulent payments, with consistent and substantial reductions of 81.25%, 84.85%, and 82.35% across the kitchen remodelling payment, laptop purchase, and HMRC tax payment scenarios, respectively.

In terms of the effects on legitimate payments, we did not find any significant differences across payment scenarios. The proportion of legitimate payments in the control group remained largely invariant, ranging from 55% to 59% across scenarios, as shown in Table 3. With regards to treatment efficacy by scenario, both the 'Risk + CTA' and 'Risk + behavioural + CTA' journeys resulted in a reduction of up to 13 percentage points in the likelihood of legitimate payments. However, the most substantial reduction was observed in the 'Control + Behavioural + CTA' payment journey, with a decrease of approximately 20 percentage points in the likelihood of legitimate payments across all payment scenarios. Despite this, even with this significant effect, the relative reduction compared to the control group was only around 36% for all payment scenarios. Therefore, our interventions resulted in a greater reduction of fraudulent payments than any unintentional effects on legitimate payments, as observed across all payment scenarios.

4.2.2 Including Demographic Controls

To mitigate potential omitted variable bias associated with demographic factors, we replicated the baseline regression analysis and included demographic controls for gender, age, and income. Table 4 presents the treatment effects accounting for these controls.

Our original conclusion that the risk-based payment journey combined with CTAs was most effective in reducing fraudulent payments held even after adding these controls, with the point estimates remaining unchanged. As in Table 1, with this payment journey, fraudulent payments declined by about 18 percentage points and legitimate payments declined by about

7 percentage points.

Turning our attention to the impact of demographic controls, the findings from Table 4 shed light on the fact that individuals aged 65 or older may be particularly less inclined to engage in both fraudulent (by 3 percentage points) and legitimate payments (by 23 percentage points) relative to those in the 18-24 age group. In part, this might result from older adults' less familiarity with modern technology, which makes them suspicious and more likely to abandon transactions at the first sign of potential fraud.

Regarding the role of income, the results suggest that income is not a significant determinant of the likelihood of either fraudulent or legitimate payments. Similarly, there were minor differences observed between genders, with women being only 3 percentage points less likely to experience fraudulent payments. It should be noted, however, that these smaller gender differences in the detection and prevention of fraud may be caused by factors other than gender itself, such as differences in financial literacy that could affect the ability to detect and prevent fraudulent activity.

4.2.3 Correcting for Multiple Hypothesis Testing

Because we have used multiple outcomes and tested various treatment effects, our next robustness test adjusts for multiple hypothesis testing (MHT) to mitigate the familywise error rates (i.e., the probability of making any type I error). We present adjusted p-values in Table 5 for fraudulent payments and in Table 6 for legitimate payments. Each table displays the baseline treatment effects, previously discussed, in Column 1, along with unadjusted p-values in Column 2. Adjusted p-values are displayed in Columns 3 to 5. Column 3 employs the correction method proposed by List et al. (2019). Bonferroni and Holm type corrections are displayed in columns 4 and 5. None of these adjustments alters our assessment of the statistical significance of the treatment effects in either fraudulent or legitimate payments.

4.3 Sensitivity Tests

4.3.1 Effects of Risk-Based Approach Accuracy

As outlined in the Experimental Design section, we divided the risk-based groups into two categories: high accuracy, where the risk-based approach correctly classified high-risk scenarios without misidentifying them as low-risk or vice versa, and low accuracy, where the approach produced some false positives by categorising legitimate scenarios as risky. Our first sensitivity test investigates whether this varying accuracy had different effects on the likelihood of engaging in fraudulent and legitimate payments. To do so, we subset the data to include all four payment journeys in the risk-based approach, and re-estimate our regressions, this time distinguishing between high and low accuracy levels by adding a dummy variable that takes the value of 1 when the risk-based approach had high accuracy, and 0 when it had low accuracy. The results are presented in Table C6 in the Appendix. The constant terms in Columns 1 and 2 represent the likelihood of fraudulent and legitimate payments, respectively, when the risk-based approach had low accuracy. We found that varying the accuracy of the risk-based approach did not have a significant effect on the likelihood of fraudulent payment. However, we did observe a slight influence on the likelihood of legitimate payments, which increased by approximately 3 percentage points when the approach had high accuracy compared to low accuracy.

4.3.2 Differential Effects of Payment Journeys Across Demographic Segments

To investigate the differential effects of our payment journeys on different demographic groups, we separately estimated the same baseline specification across various segments of the sample. Specifically, we explored differences in treatment effects across age groups, income levels, and among participants who frequently used mobile and web banking and those who did not.

Tables C7 and C8 present the results of our analysis. Table C7 focuses on fraudulent pay-

ments and shows that CTAs and behavioural messages were more effective in preventing fraud for older participants (age 55 and above) than for younger participants (18 to 34 and 35 to 54). CTAs had the largest effect, reducing fraudulent payments by approximately 16 percentage points in this age group. Although the risk-based approach alone did not reduce fraudulent payments across any age group, when combined with CTAs in the 'Risk-based + CTA' payment journey resulted in an additional 4 percentage point reduction in fraudulent payments among the older participants.

Regarding income groups, the CTA elements were most successful at preventing fraud among participants with the highest yearly income (£60,000 per annum and above). These results might reflect higher participant sophistication, as older participants and those with higher income may have greater financial literacy, greater perception of the risks associated with fraudulent activity, and therefore be more responsive to CTAs. Moreover, these groups may also be more likely to engage in a longer decision-making process and may be more willing to reconsider the payment if they have the option to cancel or save it for later.

In terms of banking habits, we found that CTAs and behavioural messages were slightly more effective in stopping fraud for participants who used web banking less frequently (less than weekly) compared to those who used it weekly. In contrast, for app banking, CTAs and behavioural messages were more effective with participants who used it at least weekly. One possible reason for this discrepancy is that participants who frequently use banking apps may have a higher level of trust in the app and be more aware of the importance of security measures, making them more responsive to CTAs, while some web users may not perceive the same level of risk and be more likely to overlook these prompts.

Regarding legitimate payments, Table C8 shows that CTA and behavioural messages reduced legitimate payments, particularly among participants with higher income. On the other hand, risk-based interventions were found to have a greater positive impact on legitimate payments for both older and lower-income groups. Other demographic segments did not

show any other substantial heterogeneous effects.

4.3.3 Effects of Scenario Order

Our next sensitivity analysis explores the effectiveness of the payment journeys based on whether participants were exposed to fraud in the first, second, or third payment scenario. The rationale behind this analysis was that individuals may become desensitised to the warnings conveyed in the app after having gone through multiple scenarios, which may reduce the warnings' efficacy.

The results in Table C9 indicate that participants were more likely to fall for fraud if they first encountered the fraudulent scenario. In the absence of any intervention, participants fell for fraud 27% of the time when their first scenario was a fraudulent request, but only 18% when it was the third.

Concerning the efficacy of the payment journeys, behavioural messages lost their effectiveness and significance when fraud appeared in the third scenario. In contrast, CTA interventions were able to maintain their significant effects, and even increase them, when participants encountered fraud in the third scenario. For instance, the 'Risk-based + CTA' payment journey resulted in an 81% reduction in fraudulent payments in the presence of fraud in the first scenario (from 27% to 5%). When fraud occurred in the third scenario, fraudulent payments decreased by 94% (from 18% to a mere 1%). However, the effects of payment journeys on legitimate payments did not demonstrate consistent patterns. Some journeys showed a greater decrease in legitimate payments in the third scenario, while others showed it in the first.

4.3.4 Additional Tests

It is important to consider the trade-off between effectiveness and user experience, as excessive warnings may lead to disengagement with the app. In additional tests, we evaluated the user experience by presenting participants with agree/disagree statements using 5-point

Likert responses. These statements assessed preferences for the app used over their current banking app, likelihood of recommending the app to friends, perceived app intuitiveness, ease of use and safety, as well as the perceived number of unnecessary steps, the frequency of reading the warnings in the app, and ease of payment cancellation. We treated the responses as binary variables (1 for "Strongly Agree" or"Agree" and 0 for other responses) and regressed them on the different payment journeys.

The results in Table C10 indicate that the 'Risk + behavioural + CTA' version of the app generally scored the highest in terms of customer satisfaction and usability metrics. Participants preferred this app over their current banking app, found it intuitive, safe, and easy to cancel payments. We also found that participants across all experimental groups were more likely to report reading the text and warnings presented in the app compared to those in the control group.

Finally, we also examined the time spent by participants in reading the scenarios, completing or canceling payments, and found no significant differences in time spent across payment journeys. On average, participants took 7 minutes per scenario (see Table C11).

5 Conclusion

APP fraud is a new form of fraud, facilitated by innovations in payments technology, which is costly to consumers. Our study sheds light on the effectiveness of various interventions in preventing APP fraud in online transactions. Our interventions demonstrated economically large and statistically significant effects on the likelihood to fall for fraud. The most significant effects were observed when altering the CTAs presented in the banking app to allow for greater payment cancellation and deferral options, in conjunction with the provision of warnings for high-risk payments. These modifications caused a substantial reduction in the likelihood of falling for fraudulent activity, with participants being 81% less likely to fall for fraud than those in the control group. CTAs were especially effective for older participants

and those with higher yearly income. In comparison, behavioural interventions had smaller effects, reducing the likelihood of falling for fraud by only 18%.

We also found that the efficacy of interventions is contingent on the underlying psychological factors involved in each particular scam. For example, messages that leverage loss aversion were more effective in preventing purchase scams on Facebook Marketplace, while they were less effective in impersonation scams involving overdue taxes to HMRC, likely because the fear of legal consequences might have outweighed the impact of loss aversion on participants.

Furthermore, our findings indicate that participants may develop a tolerance to certain interventions over time and become less responsive to them. In particular, behavioural messages lost effectiveness over time. In contrast, CTA interventions were able to maintain their effectiveness and even strengthen when participants encountered fraud in the third scenario, leading to a 94% reduction in fraudulent payments.

When deciding which intervention to recommend, we should consider the trade-off between combating fraudulent payments and preserving legitimate transactions. While offering participants more opportunities to cancel and defer payments came at a cost of dissuading them from making slightly suspicious looking, albeit legitimate, payments, we found that the magnitude of the reduction in fraudulent payments was much larger than the unintended effect on legitimate payments. For instance, the 'Risk-based + CTA' journey reduced the percentage of fraudulent payments by 81%, while legitimate payments only decreased by 12%. We should note that this slight reduction in legitimate payments may result from participants becoming more cautious when presented with payment options. However, this cautious behaviour can lead to a more informed and thoughtful decision-making process in the future, ultimately benefiting them.

Overall, our study provides empirical evidence of the effectiveness of various fraud prevention measures, highlighting the potential importance of CTAs in reducing fraud. As the number of consumers engaging in online transactions continues to rise, the risk of fraud also increases.

Our study contributes to the limited empirical evidence on the effectiveness of different fraud prevention measures and can guide policymakers and businesses in developing effective strategies to prevent fraud in online transactions.

References

Allen, F., X. Gu, and J. Jagtiani (2021). A survey of fintech research and policy discussion. *Review of Corporate Finance 1*, 259–339.

Berg, T., A. Fuster, and M. Puri (2022). Fintech lending. *Annual Review of Financial Economics 14*, 187–207.

Buchak, G., J. Hu, and S.-J. Wei (2021). Fintech as a financial liberator. Technical report, National Bureau of Economic Research.

Button, M. and C. Cross (2017). Technology and fraud: The 'fraudogenic'consequences of the internet revolution. In *The Routledge handbook of technology, crime and justice*, pp. 78–95. Routledge.

Chang, J. J. and M. D. Chong (2010). Psychological influences in e-mail fraud. *Journal of Financial Crime*.

Chen, H., C. E. Beaudoin, and T. Hong (2017). Securing online privacy: An empirical test on internet scam victimization, online privacy concerns, and privacy protection behaviors. *Computers in human behavior 70*, 291–302.

Dahlgreen, J. (2021). Catastrophic fraud loss lies where it falls? push payment scams and the bank's duty of care to its customer. *Journal of Financial Crime*.

Di Maggio, M., D. Ratnadiwakara, and D. Carmichael (2022). Invisible primes: Fintech lending with alternative data. Technical report, National Bureau of Economic Research.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: personality and cognitive correlates. *Journal of personality and social psychology 58*(1), 95.

Finn, P. and M. Jakobsson (2007). Designing ethical phishing experiments. *IEEE Technology and Society Magazine 26*(1), 46–58.

Griffin, J. M., S. Kruger, and P. Mahajan (2023). Did fintech lenders facilitate ppp fraud? *The Journal of Finance*.

Hayashi, F. (2012). Mobile payments: What's in it for consumers? *Economic Review-Federal Reserve Bank of Kansas City*, 35.

He, Z., J. Huang, and J. Zhou (2023). Open banking: Credit market competition when borrowers own the data. *Journal of financial economics 147*(2), 449–474.

Johnson, M. J., I. Ben-David, J. Lee, and V. Yao (2023). Fintech lending with lowtech pricing. Technical report, National Bureau of Economic Research.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*(4), 773–793.

Luo, X. R., W. Zhang, S. Burd, and A. Seazzu (2013). Investigating phishing victimization with the heuristic–systematic model: A theoretical framework and an exploration. *Computers & Security 38*, 28–38.

Ma, K. W. F., T. Dhot, and M. Raza (2023). Considerations for using artificial intelligence to manage authorized push payment (app) scams. *IEEE Engineering Management Review*.

Maher, R. (2021). A critical analysis of recent efforts in the united kingdom to tackle authorised push payment scams and the impact on the bank-customer relationship. *Trinity CL Rev. 24*, 134.

McIlroy, D. and R. Sethi-Smith (2021). Prospects for bankers' liability for authorised push payment fraud. *Butterworths Journal of International Banking and Financial Law*, 1–1.

Parlour, C. A., U. Rajan, and H. Zhu (2022). When fintech competes for payment flows. *The Review of Financial Studies 35*(11), 4985–5024.

Pattinson, M. R., C. Jerram, K. Parsons, A. McCormac, and M. A. Butavicius (2011). Managing phishing emails: A scenario-based experiment. In *HAISA*, pp. 74–85.

Petty, R. E., J. T. Cacioppo, R. E. Petty, and J. T. Cacioppo (1986). *The elaboration likelihood model of persuasion*. Springer.

Philippon, T. (2016). The fintech opportunity. Technical report, National Bureau of Economic Research.

PSR (2023). Annual fraud report 2023. *Payment Systems Regulator*.

Reisig, M. D. and K. Holtfreter (2013). Shopping fraud victimization among the elderly. *Journal of Financial Crime*.

Taylor, J. L. and T. Galica (2020). A new code to protect victims in the uk from authorised push payments fraud. *Banking & Finance Law Review 35*(2), 327–332.

Wilsem, J. v. (2013). Hacking and harassment—do they have something in common? comparing risk factors for online victimization. *Journal of Contemporary Criminal Justice 29*(4), 437–453.

Yang, T. and X. Zhang (2022). Fintech adoption and financial inclusion: Evidence from household consumption in china. *Journal of Banking & Finance 145*, 106668.

Ziegler, T. (2021). Implementation of open banking protocols around the world. *The Palgrave Handbook of Technological Finance*, 751–779.

Figure 1: Detail of Purchase Scam Scenario

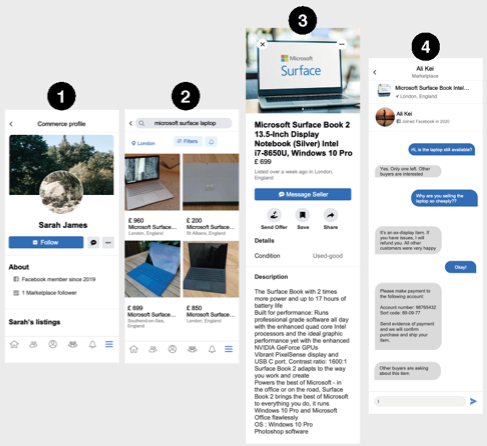
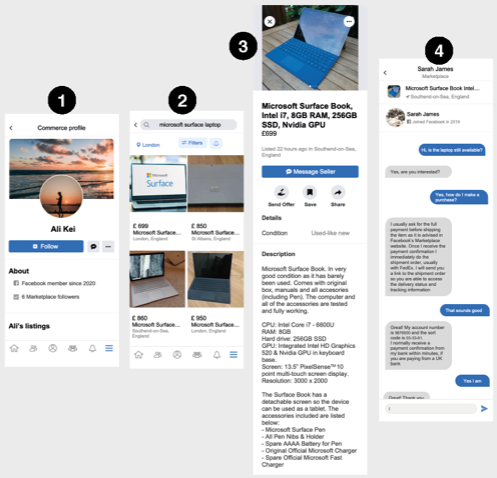
Scenario	Purchase scam	
	Legitimate	**Illegitimate**
Scenario description (shared with participamts)	The laptop you use for work has recently broken down and you must replace it urgently. You had paid £1,499.99 for your old computer, a Microsoft Surface Laptop. You have been searching on Facebook Marketplace for the same model. You find a model you like from a seller that you recognize. Your close friend has purchased from this seller before. Please review the following screens and proceed with the payment as you see fit. Making the purchase will earn you 50% of the value of the payment. If you do not make the purchase you may incur a loss of £1,000 (from not being able to work).	The laptop you use for work has recently broken down and you must replace it urgently. You had paid £1,499.99 for your old computer, a Microsoft Surface Laptop. You have been searching on Facebook Marketplace for the same model. You find a model you like from a seller in an online Marketplace. Please review the following screens and proceed with the payment as you see fit. Making the purchase will earn you 50% of the value of the payment. If you do not make the purchase you may incur a loss of £1,000 (from not being able to work).
Proof		

Figure 2: Visualisation of the experiment design

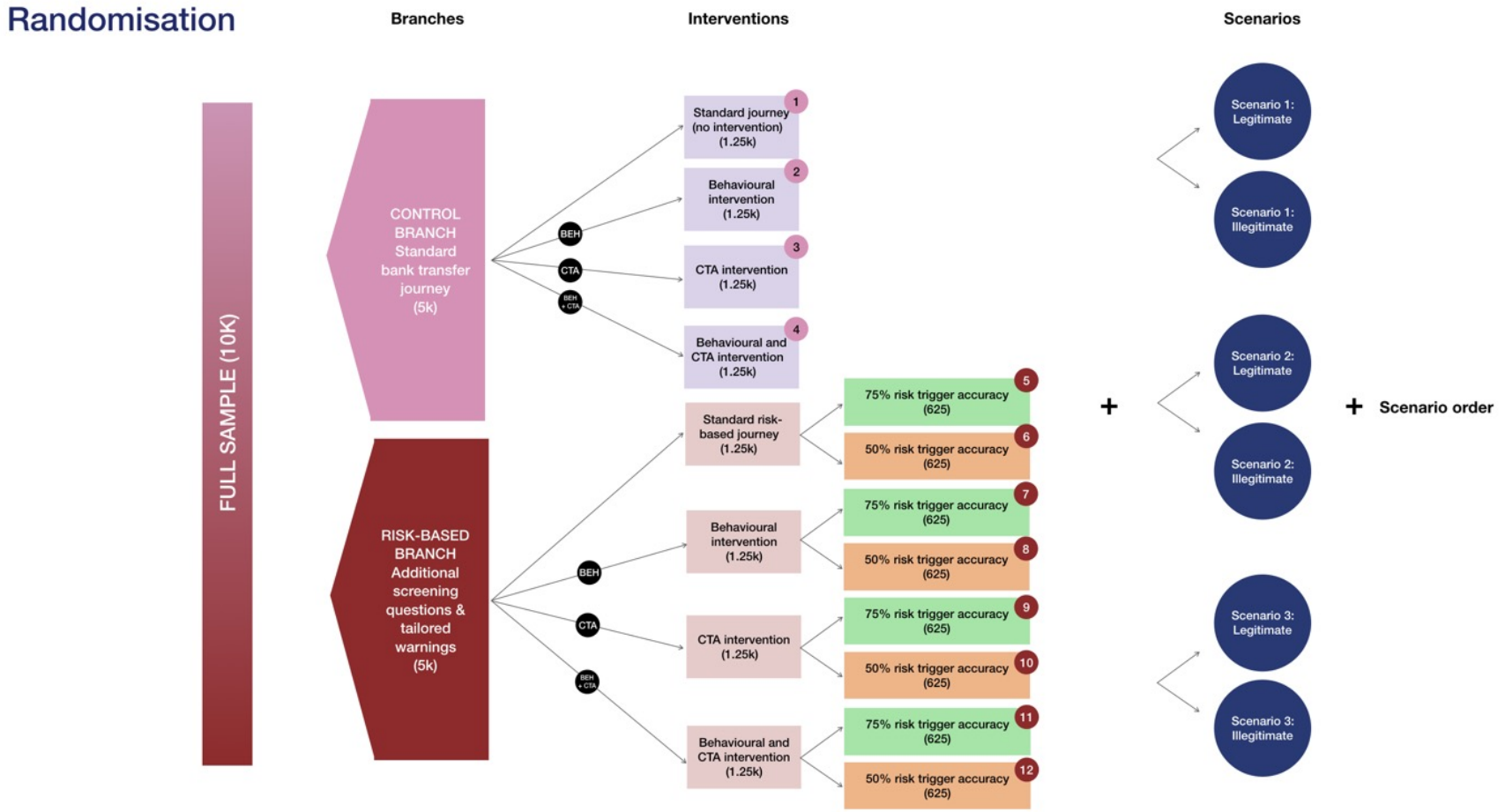


Figure 3: Excerpts from the Control-Behavioural journey

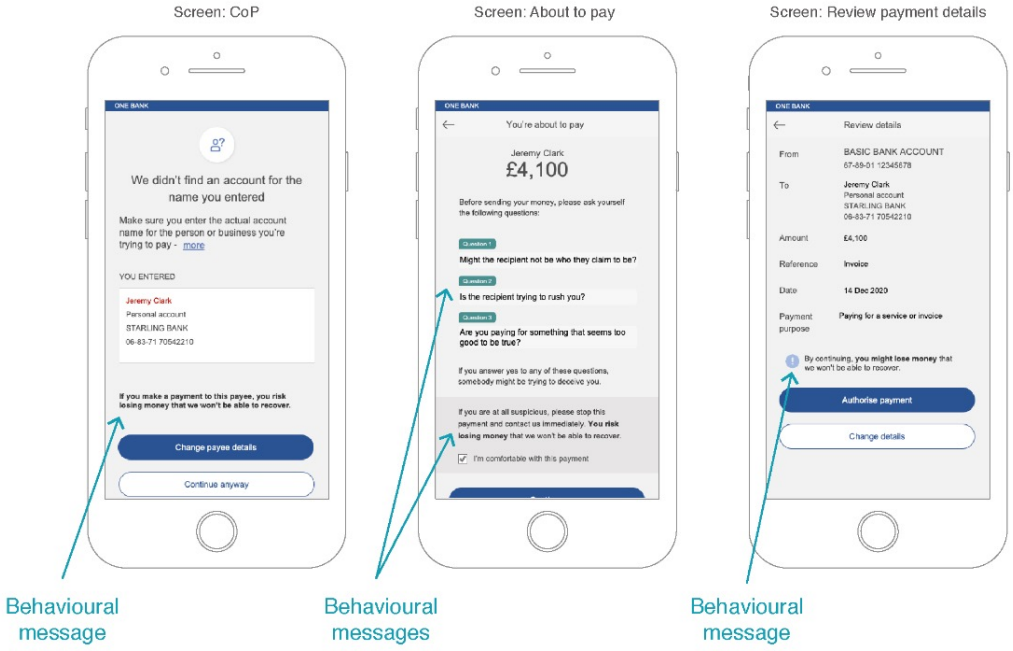


Figure 4: Excerpts from the CTA intervention (control branch)

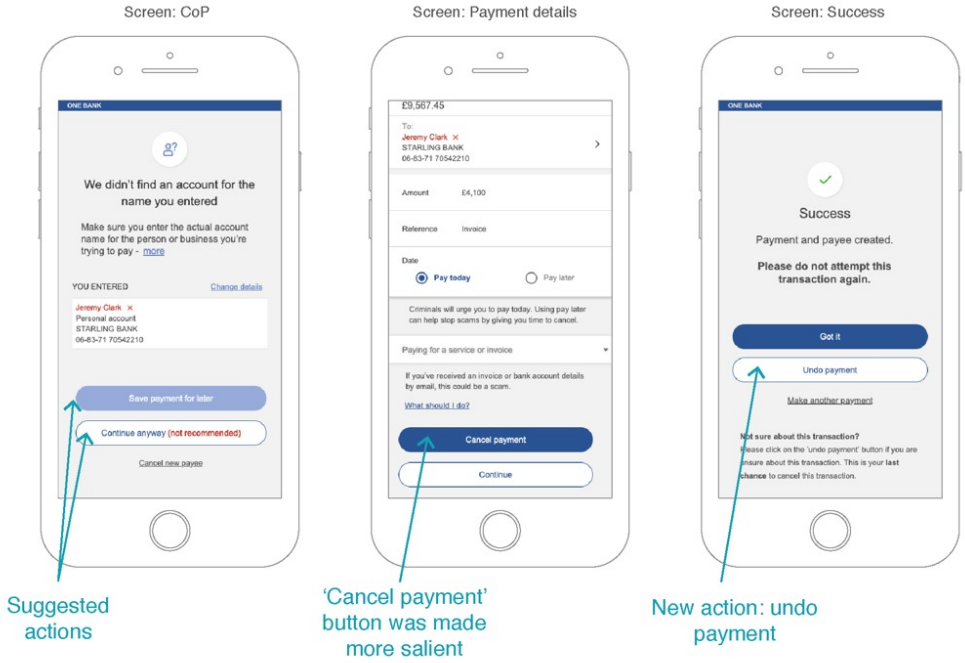


Figure 5: Excerpts from the Behavioural-CTA intervention (control branch)

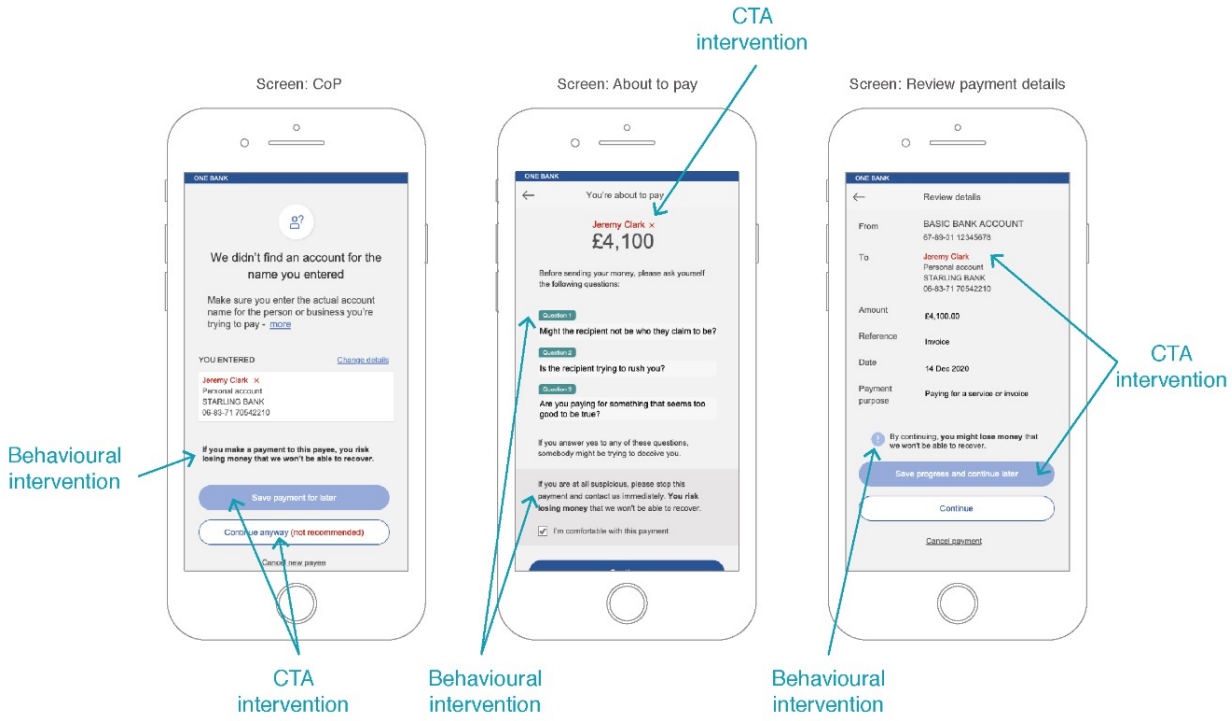


Figure 6: Additional screening questions (risk-based branch)

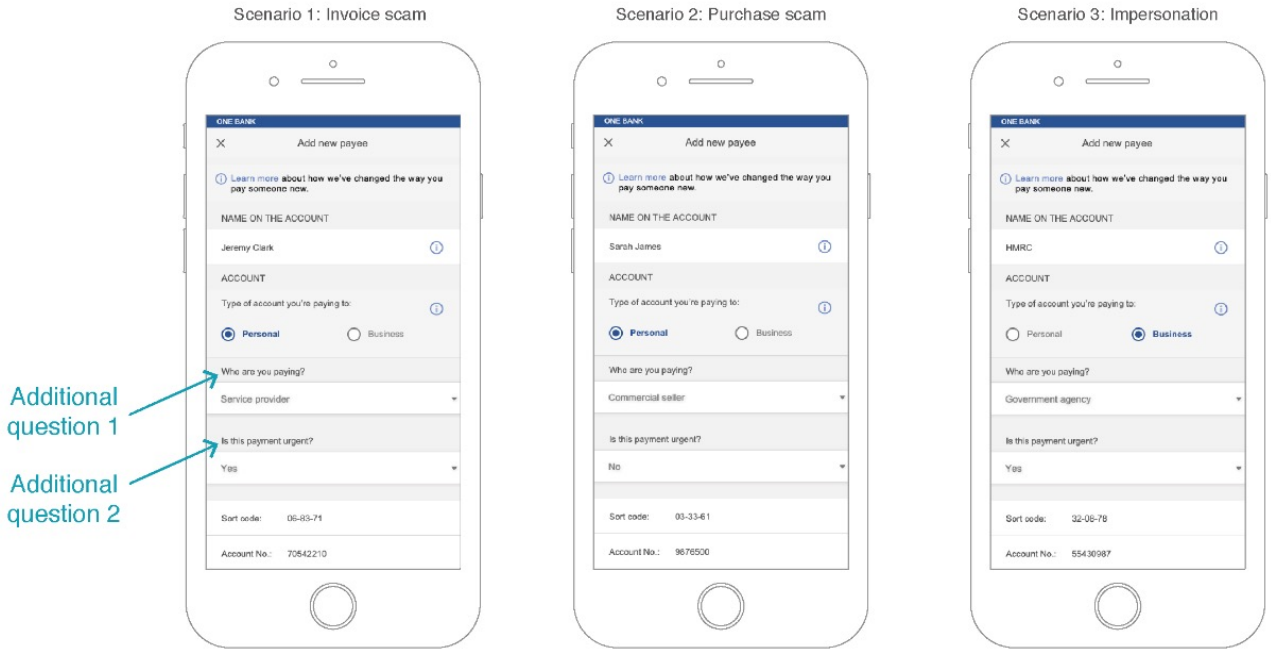


Figure 7: Excerpts from the Behavioural intervention (risk-based branch)

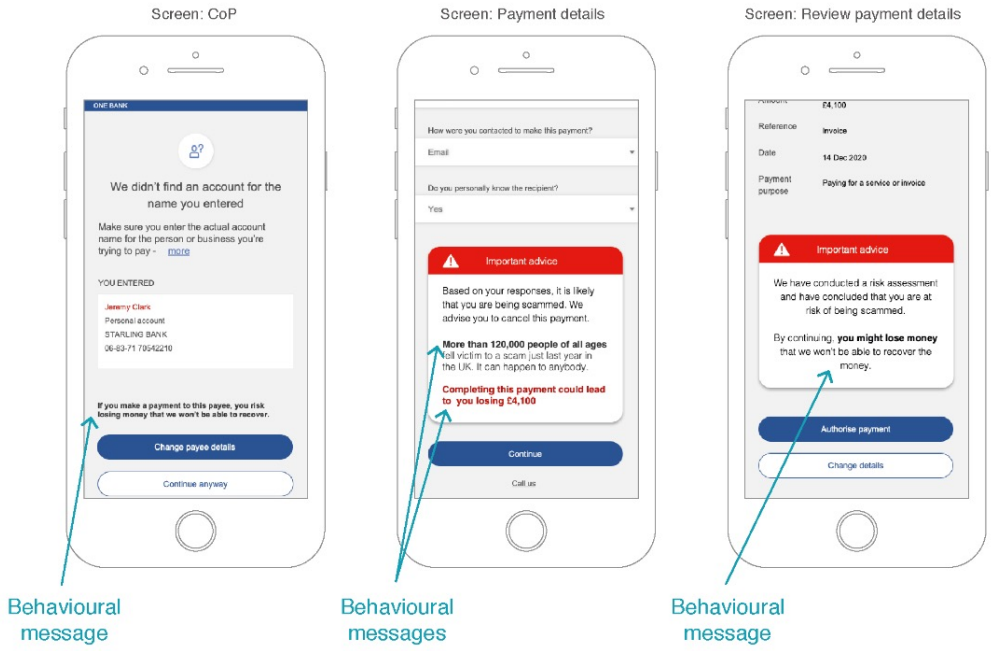


Figure 8: Excerpts from the CTA intervention (risk-based branch)

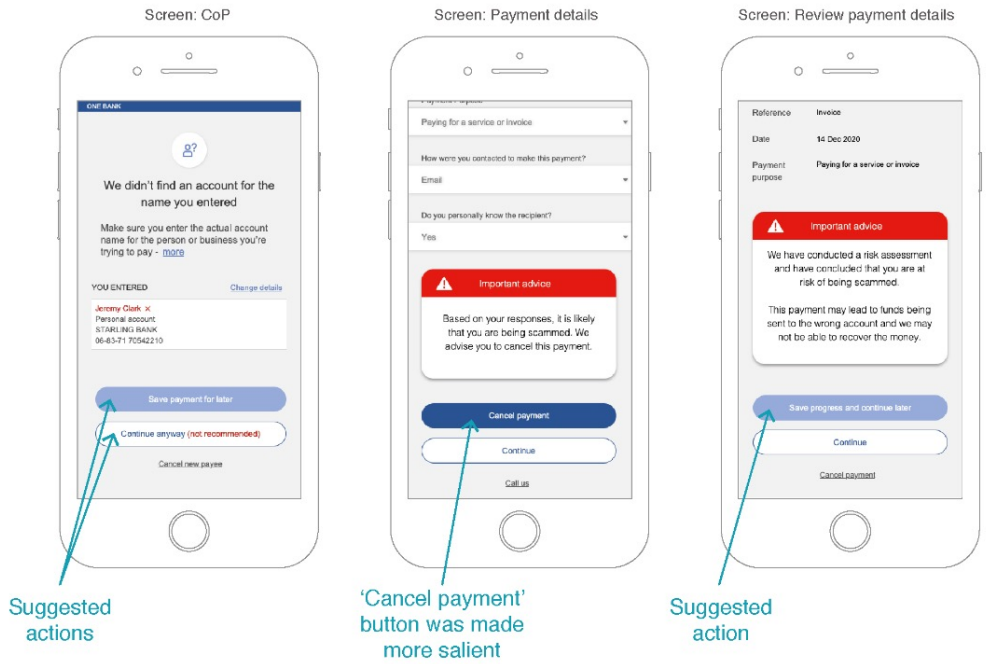


Figure 9: Excerpts from the Behavioural-CTA intervention (risk-based branch)

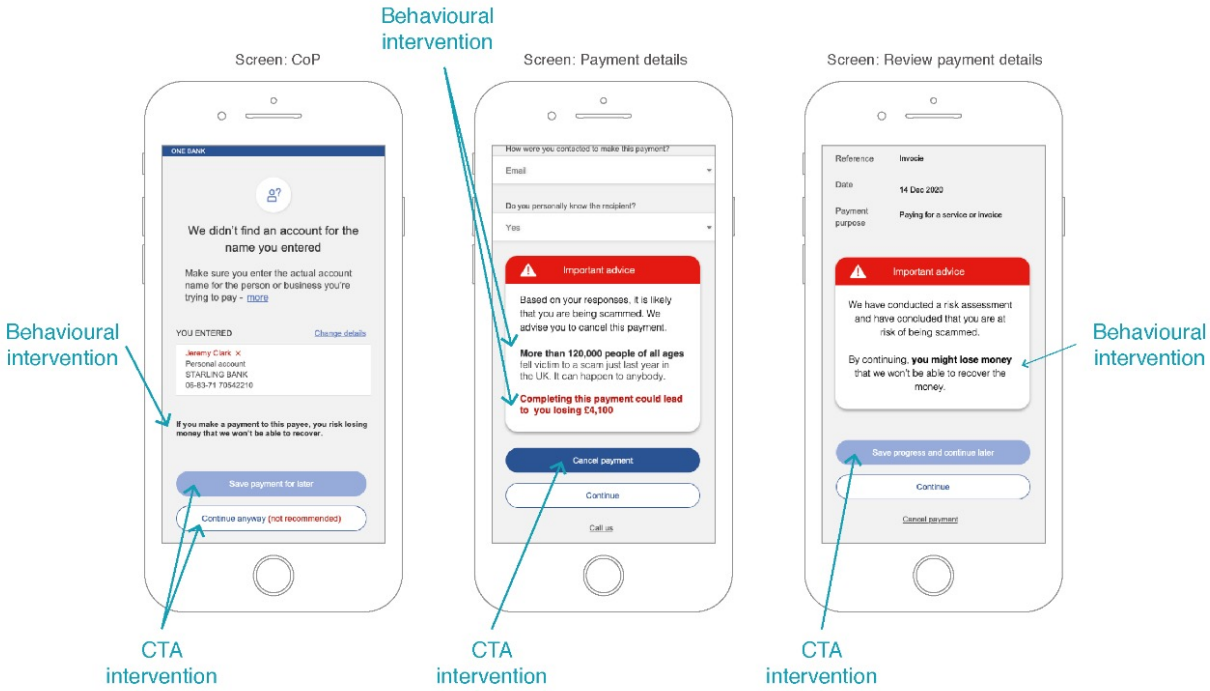


Figure 10: Treatment effects on payment behaviour

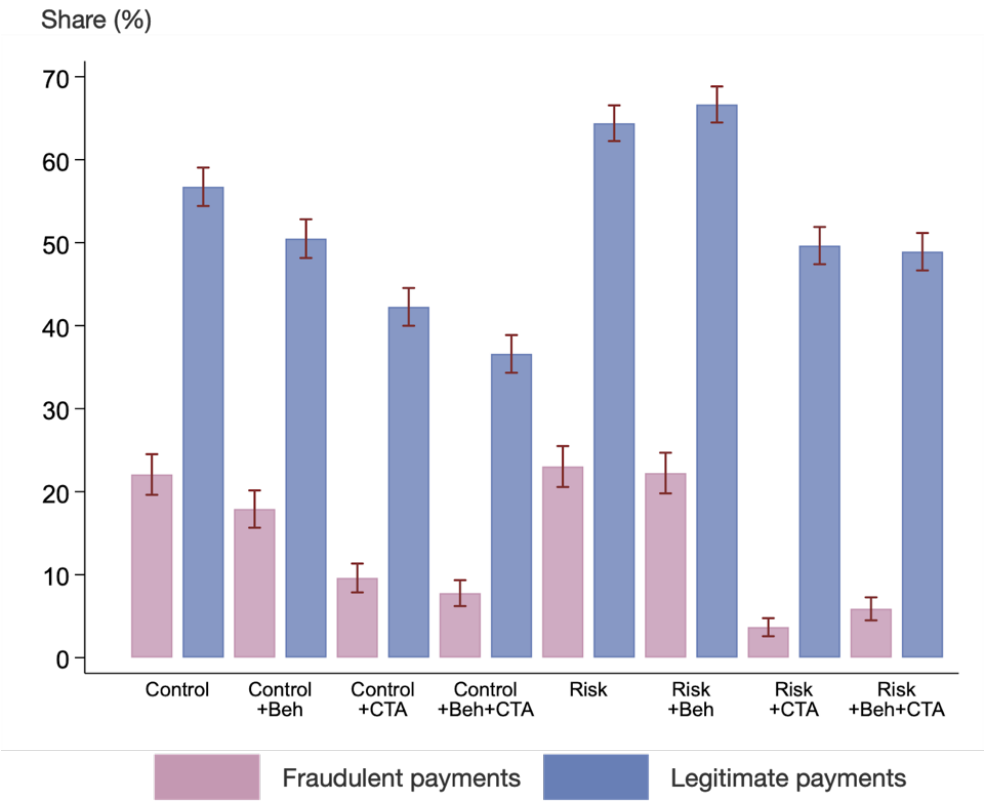


Table 1: Treatment effects on payment behaviour

	% made a fraudulent payment (1)	% made a legitimate payment (2)
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	0.00	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)
Constant	0.22***	0.57***
	(0.01)	(0.01)
Observations	8958	8958
R-squared	0.048	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). Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table 2: Treatment effects on fraudulent payments by scenario

	% made fraudulent kitchen remodelling payment	% made fraudulent laptop purchase	% made fraudulent HMRC tax payment
	(1)	(2)	(3)
Control + behavioural	-0.04*	-0.08**	-0.01
	(0.03)	(0.03)	(0.03)
Control + CTA	-0.10***	-0.16***	-0.12***
	(0.02)	(0.03)	(0.02)
Control + behavioural + CTA	-0.09***	-0.22***	-0.12***
	(0.02)	(0.03)	(0.02)
Risk-based	-0.04	0.04	0.03
	(0.03)	(0.04)	(0.03)
Risk-based + behavioural	-0.06**	0.08**	-0.01
	(0.02)	(0.04)	(0.03)
Risk-based + CTA	-0.13***	-0.28***	-0.14***
	(0.02)	(0.03)	(0.02)
Risk-based + behavioural + CTA	-0.11***	-0.28***	-0.09***
	(0.02)	(0.03)	(0.02)
Constant	0.16***	0.33***	0.17***
	(0.02)	(0.02)	(0.02)
Observations	2987	3000	2971
R-squared	0.021	0.103	0.040

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). Columns 1 to 3 are restricted to those who were shown a fraudulent version of scenarios 1, 2 or 3, respectively. Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table 3: Treatment effects on legitimate payments by scenario

	% made legitimate kitchen remodelling payment	% made legitimate laptop purchase	% made legitimate HMRC tax payment
	(1)	(2)	(3)
Control + behavioural	-0.08***	-0.06*	-0.05*
	(0.03)	(0.03)	(0.03)
Control + CTA	-0.11***	-0.13***	-0.19***
	(0.03)	(0.03)	(0.03)
Control+ behavioural + CTA	-0.20***	-0.21***	-0.20***
	(0.03)	(0.03)	(0.03)
Risk-based	0.15***	0.02	0.07**
	(0.03)	(0.03)	(0.03)
Risk-based + behavioural	0.15***	0.06**	0.09***
	(0.03)	(0.03)	(0.03)
Risk-based + CTA	0.05*	-0.13***	-0.13***
	(0.03)	(0.03)	(0.03)
Risk-based + behavioural + CTA	-0.04	-0.10***	-0.10***
	(0.03)	(0.03)	(0.03)
Constant	0.59***	0.57***	0.55***
	(0.02)	(0.02)	(0.02)
Observations	2987	3000	2971
R-squared	0.077	0.048	0.071

Notes: The regressions were conducted using a Linear Probability Model (LPM). The outcomes are the share of legitimate payments made per participant (semi-continuous variable that can take the values 0, 0.5, or 1). Columns 1 to 3 are restricted to those who were shown a fraudulent version of scenarios 1, 2 or 3, respectively. Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table 4: Treatment effects on payment behaviour

	% made a fraudulent payment (1)	% made a legitimate payment (2)
Control + behavioural	-0.04***	-0.06***
	(0.01)	(0.02)
Control + CTA	-0.12***	-0.15***
	(0.01)	(0.02)
Control + behavioural + CTA	-0.14***	-0.21***
	(0.01)	(0.02)
Risk-based	0.01	0.07***
	(0.01)	(0.02)
Risk-based + behavioural	0.00	0.10***
	(0.01)	(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)
Female	-0.02***	-0.06***
	(0.01)	(0.01)
Age (reference 18-24)		
Age 25-34	0.03**	-0.01
	(0.01)	(0.01)
Age 35-44	0.01	-0.05***
	(0.01)	(0.01)
Age 45-54	-0.01	-0.12***
	(0.01)	(0.01)
Age 55-64	-0.01	-0.17***
	(0.01)	(0.02)
Age 65+	-0.03**	-0.23***
	(0.02)	(0.02)
Income (reference £20k or less)		
£20k-£40k	-0.01	0.01
	(0.01)	(0.01)
£40k-£60k	-0.02*	0.02
	(0.01)	(0.01)
£60k +	-0.00	0.01
	(0.02)	(0.02)
Constant	0.24***	0.68***
	(0.01)	(0.02)
Observations	8958	8958
R-squared	0.051	0.098

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). Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table 5: Treatment effects on fraudulent payments - Corrections for multiple hypothesis testing

	Treatment effect	p values			
		Unadjusted	MHT	Bonferroni	Holm
	(1)	(2)	(3)	(4)	(5)
Control + behavioural	-0.0416	0.0147	0.0390	0.2053	0.0440
Control + CTA	-0.1247	0.0003	0.0003	0.0047	0.0037
Control + behavioural + CTA	-0.1430	0.0003	0.0003	0.0047	0.0017
Risk-based	0.0096	0.5903	0.8137	1.0000	1.0000
Risk-based + behavioural	0.0017	0.9197	0.9197	1.0000	0.9197
Risk-based + CTA	-0.1840	0.0003	0.0003	0.0047	0.0040
Risk-based + behavioural + CTA	-0.1619	0.0003	0.0003	0.0047	0.0013

Notes: The table shows the treatment effects on fraudulent payments an adjusted p-values corrected for multiple hypothesis testing. MHT in Column 3 uses the correction proposed by List et al. (2019) to reduce the familywise error rates for multiple hypothesis tests (i.e., the probability of making any type I error). Bonferroni or Holm type corrections are displayed in Columns 4 and 5. All p-values are calculated using the Stata package mhtexp with 3,000 bootstrap replications (List et al., 2019).

Table 6: Treatment effects on legitimate payments - Corrections for multiple hypothesis testing

	Treatment effect	p values			
		Unadjusted	MHT	Bonferroni	Holm
	(1)	(2)	(3)	(4)	(5)
Control + behavioural	-0.0625	0.0003	0.0003	0.0047	0.0033
Control + CTA	-0.1447	0.0003	0.0003	0.0047	0.0043
Control + behavioural + CTA	-0.2014	0.0003	0.0003	0.0047	0.0047
Risk-based	0.0766	0.0003	0.0003	0.0047	0.002
Risk-based + behavioural	0.0992	0.0003	0.0003	0.0047	0.003
Risk-based + CTA	-0.0708	0.0003	0.0003	0.0047	0.0027
Risk-based + behavioural + CTA	-0.0782	0.0003	0.0003	0.0047	0.0023

Notes: The table shows the treatment effects on legitimate payments an adjusted p-values corrected for multiple hypothesis testing. MHT in Column 3 uses the correction proposed by List et al. (2019) to reduce the familywise error rates for multiple hypothesis tests (i.e., the probability of making any type I error). Bonferroni or Holm type corrections are displayed in Columns 4 and 5. All p-values are calculated using the Stata package mhtexp with 3,000 bootstrap replications (List et al., 2019).

Appendix A: Scenarios Used in the Experiment

Figure A1: Detail of Invoice Scam Scenario

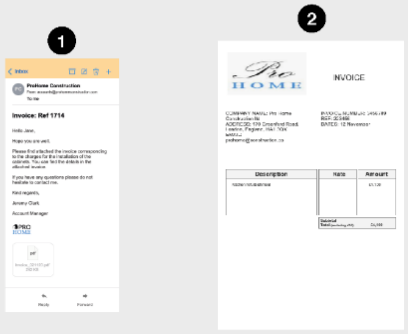
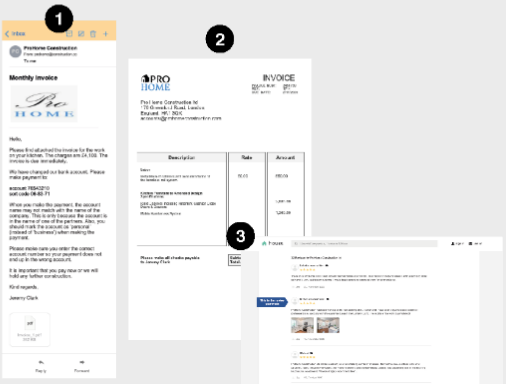
Scenario	Invoice scam	
	Legitimate	Illegitimate
Scenario description (shared with participamts)	You have been renovating the kitchen of your house. This has involved everything from redoing the sink and plumbing fixtures, to changing the flooring and getting new appliances, cabinets and countertops. You have hired ProHome Construction as the contractor to complete the works. Over the course of the project, you have been corresponding with John Smith (the account manager) at the email address accounts@prohomeconstruction.com. The company has already completed the work on the sink and plumbing and are due to install new cabinets this week. You are expecting an invoice from the company. Please review the following screens and proceed with the payment as you see fit. Making the payment in time will earn you 50% of the value of the payment. If you do not pay, you may incur a loss of £1,000 (from delaying the project).	
Proof		

Figure A2: Detail of Impersonation Scam Scenario

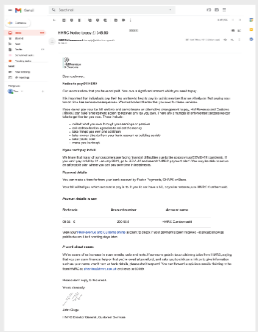
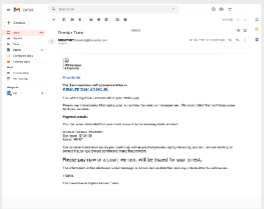
Scenario	Impersonation scam	
	Legitimate	Illegitimate
Scenario description (shared with participants)	You usually file Self Assessment tax returns with HMRC but have not yet paid your tax this year. You just received an email explaining that you have not paid the balance due (£1345.90), with instructions 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).	
Proof	1 	1 

Table A1: Types of APP scams

Types of APP scams	Comm. channels	Volume (share)	Avg. loss (share)	Overview
Invoice & mandate scams	Email	7k (9%)	£16k (35%)	Victim is intercepted with a a request to make a payment to a different account
Impersonation: bank staff/police	Phone/SMS	5k (6%)	£10k (16%)	Bank staff/police urges transfer to 'safe' account
Impersonation: other	Phone/SMS	5k (6%)	£6k (10%)	Victim is asked to pay overdue tax or fee (may be gov or utilities)
CEO fraud	Phone/Email	600 (1%)	£24k (4%)	Impersonating the victim's CEO and asking for urgent payment
Purchase scam	Ecommerce	56k (64%)	£800 (13%)	Victim pays for what seem legitimate goods/services, but they are never delivered
Investment scam	Online ad	3k (4%)	£14k (14%)	Victim is invited to a fictitious investment scheme
Advance fee scam	Online ad	8k (9%)	£2k (4%)	Victim is asked to make a small payment to receive a larger sum
Romance scam	Social media	2k (2%)	£9k (4%)	Emergency request after romantic relationship is established

Source: UK Finance (2020)

Table A2: Tell-tale signs of scam scenarios used in experiment

Invoice & Mandate Scams

- Name/address of sender does not match exactly other emails from the same party.
- The email purports to be a "confidential" or "private" request.
- An email contains an attachment that purports to be an order confirmation or receipt.
- The sender's email address does not seem to match the contents.
- The wording of the email is awkward.
- Logo in the email or invoice may be the same as the authentic one but is blurred (might be scanned).

Purchase Scams

- The sender's email address does not seem to match the contents.
- The wording of the email is awkward.
- They will request payment using a preloaded money card or bank transfer.
- The prices of the products will be much lower than other shops and sellers will urge victims to buy quickly as the sale is a limited time offer or in high demand.
- They will only show a post office box rather than a full postal address.
- The URL contains spelling mistakes in the shop name.
- Missing terms and conditions, or delivery information. Seller or store does not appear on Google searches.

Impersonation Scams

- The email threatens the victim with dire consequences if they do not comply.
- The email asks for "urgent" or "immediate" action.
- The email purports to be a "confidential" or "private" request.
- The email has an attachment with some non-standard document extension.
- During tax season there is a bump in spear phishing and telephone scams by "tax authorities" requesting financial information or providing tax "receipts" that are malware in disguise.
- The sender's email address does not seem to match the contents.
- The wording of the email is awkward.

Appendix B: Payment Journeys

Figure B1: Payment task 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.

To make 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).

Figure B2: Screens most representative of each treatment group. Control branch.

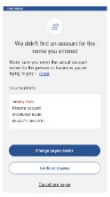
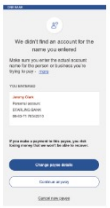
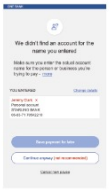
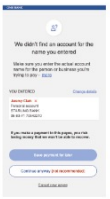
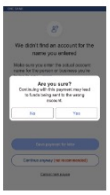
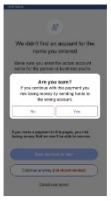
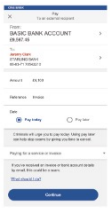
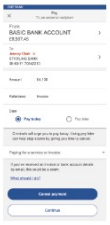
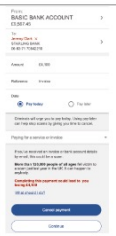
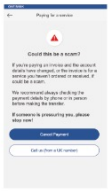
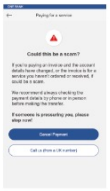
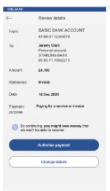
Screen	Group 1 Control	Group 2 Behavioural	Group 3 CTA	Group 4 Behavioural-CTA
CoP				
CoP pop-up				
Payment details				
CRM				
About to pay				
Review payment				

Figure B3: Screens most representative of each treatment group. Risk based branch.

Figure B4: User journey of the control group. Control branch.

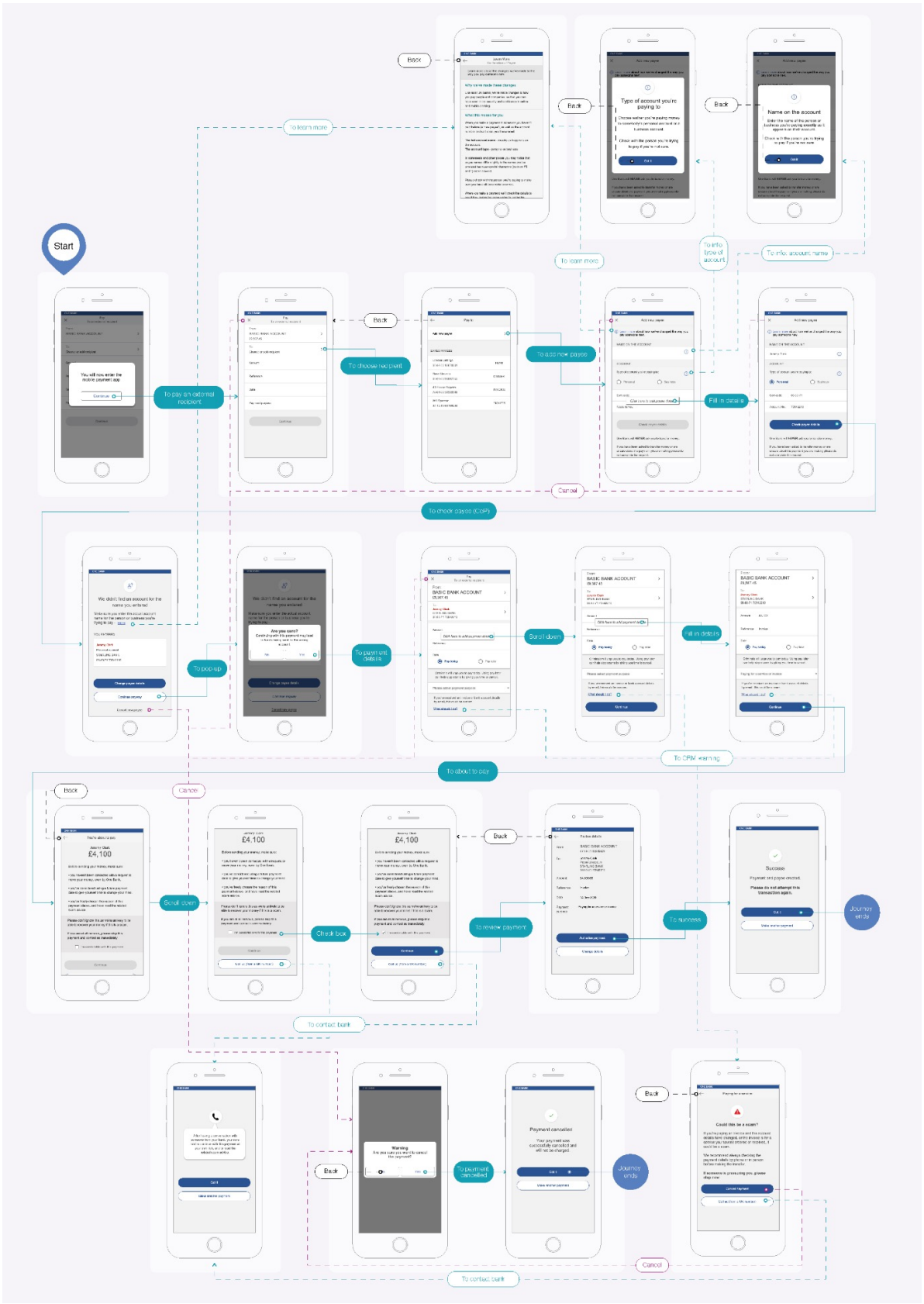
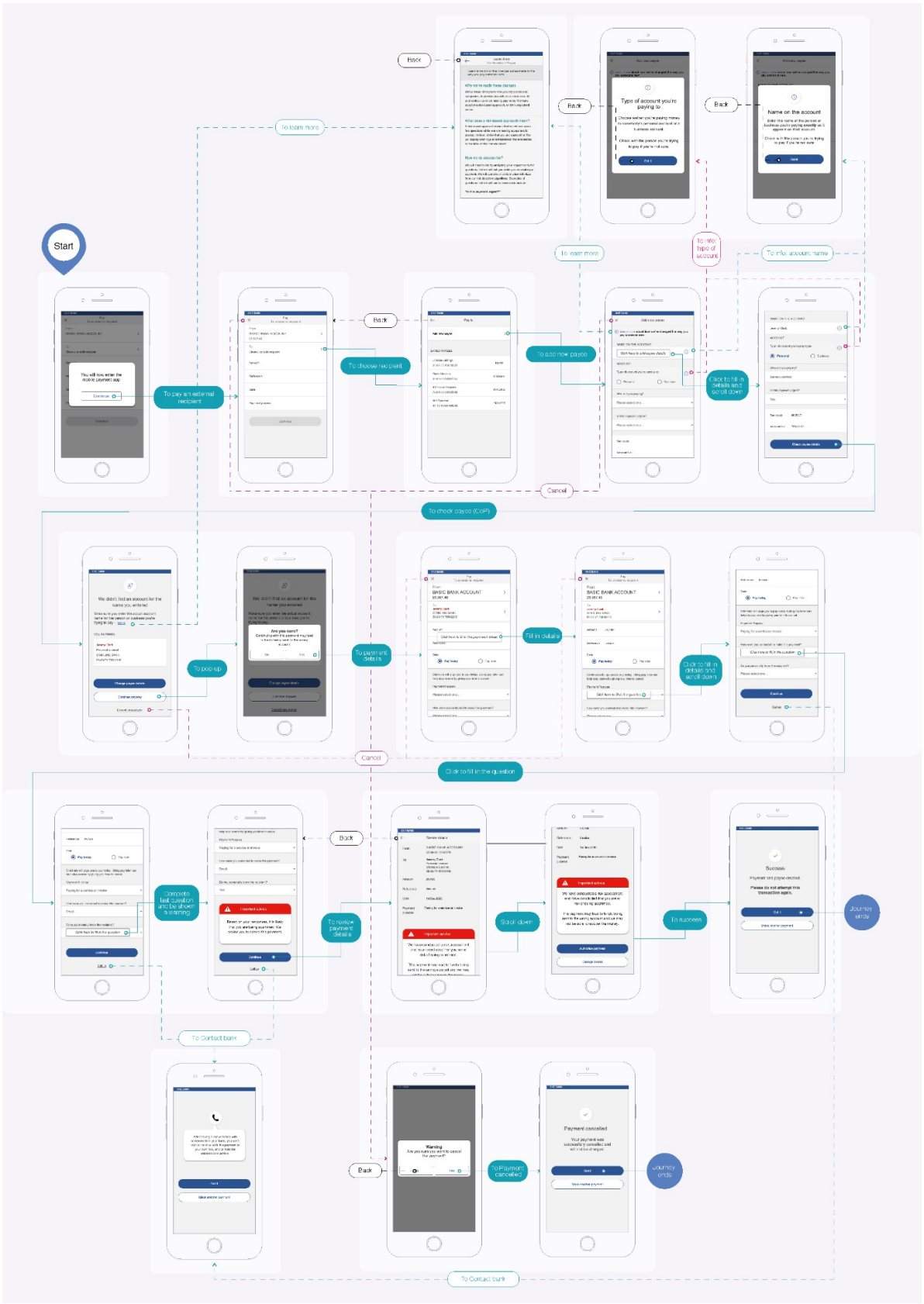


Figure B5: User journey of the control group. Risk based branch.



Appendix C: Additional Analyses

Table C1: Characteristics of participants

Characteristic	Sample
Female	54%
Male	46%
Age: 18-24	14%
Age: 25-34	24%
Age: 35-44	19%
Age: 45-54	19%
Age: 55-64	15%
Age: 65+	10%
Region: East Midlands	7%
Region: East of England	8%
Region: Greater London	13%
Region: North East	4%
Region: North West	12%
Region: Northern Ireland	2%
Region: Scotland	9%
Region: South East	15%
Region: South West	8%
Region: Wales	4%
Region: West Midlands	9%
Region: Yorkshire and the Humber	9%
Income: £20k or less	42%
Income: £20-£40k	40%
Income: £40-£60k	12%
Income: £60k+	6%
Use app banking on a weekly basis	74%
Use web banking on a weekly basis	55%

Table C2: Balance across banking journeys

Characteristic	G1	G2	G3	G4	G5	G6	G7	G8	p-value
Female	55%	57%	54%	53%	53%	55%	53%	52%	0.218
Male	45%	43%	45%	47%	46%	45%	47%	48%	0.212
Age:									
18-24	15%	13%	13%	14%	14%	13%	14%	13%	0.949
25-34	22%	25%	24%	25%	22%	24%	23%	23%	0.572
35-44	19%	20%	19%	19%	20%	19%	18%	19%	0.856
45-54	18%	17%	20%	19%	19%	17%	21%	20%	0.144
55-64	16%	15%	15%	13%	15%	17%	15%	15%	0.606
65+	11%	10%	10%	10%	10%	10%	10%	10%	0.982
Region:									
East Midlands	6%	8%	8%	8%	6%	8%	8%	8%	0.308
East of England	8%	8%	7%	8%	8%	7%	6%	10%	0.076
Greater London	12%	13%	15%	15%	14%	13%	13%	13%	0.636
North East	5%	4%	4%	4%	5%	5%	5%	4%	0.811
North West	12%	11%	12%	11%	12%	12%	11%	10%	0.751
Northern Ireland	2%	2%	2%	2%	2%	2%	2%	2%	0.862
Scotland	8%	10%	7%	9%	10%	10%	10%	10%	0.342
South East	15%	14%	15%	14%	15%	16%	16%	15%	0.721
South West	10%	8%	8%	7%	9%	8%	8%	8%	0.419
Wales	5%	4%	4%	4%	5%	3%	5%	4%	0.551
West Midlands	9%	9%	9%	9%	7%	10%	9%	9%	0.624
Yorkshire and the Humber	8%	9%	9%	10%	9%	8%	9%	9%	0.886
Income:									
£20k or less	42%	41%	39%	42%	44%	45%	43%	41%	0.165
£20-£40k	39%	42%	41%	39%	39%	37%	41%	40%	0.304
£40-£60k	12%	12%	13%	12%	11%	12%	11%	13%	0.756
£60k+	7%	5%	7%	7%	6%	6%	6%	7%	0.452
Weekly use bank app	75%	73%	74%	76%	75%	72%	75%	76%	0.326
Daily use bank app	55%	54%	57%	54%	58%	55%	54%	56%	0.534
N	1106	1123	1105	1134	1125	1111	1147	1107	

Notes: P-values are from joint-orthogonality tests across groups. Journey groups refer to: G1 - Control; G2 - Control + Behavioural; G3 - Control + CTA; G4 - Control + Behavioural + CTA; G5 - Risk Based; G6 - Risk Based + Behavioural ; G7 - Risk Based + CTA ; G8 - Risk Based + Behavioural + CTA

Table C3: Balance across assigned scam type

Characteristic	Kitchen remodelling	Laptop purchase	HMRC overdue taxes	p-value
Female	54%	53%	55%	0.216
Male	46%	46%	44%	0.232
Age:				
18-24	13%	14%	14%	0.250
25-34	24%	23%	24%	0.587
35-44	19%	19%	20%	0.819
45-54	19%	19%	18%	0.478
55-64	16%	16%	14%	0.110
65+	10%	9%	10%	0.460
Region:				
East Midlands	8%	8%	7%	0.257
East of England	8%	7%	8%	0.273
Greater London	14%	14%	13%	0.543
North East	4%	5%	4%	0.161
North West	11%	12%	11%	0.814
Northern Ireland	2%	2%	2%	0.728
Scotland	8%	9%	10%	0.010
South East	14%	15%	16%	0.330
South West	8%	8%	9%	0.445
Wales	5%	4%	4%	0.070
West Midlands	9%	9%	8%	0.472
Yorkshire and the Humber	9%	9%	8%	0.218
Income:				
£20k or less	41%	42%	43%	0.579
£20-£40k	41%	39%	39%	0.285
£40-£60k	12%	12%	11%	0.544
£60k+	6%	7%	7%	0.067
Weekly use bank app	74%	75%	75%	0.750
Daily use bank app	56%	55%	55%	0.940
N	2987	3000	2971	

Notes: P-values are from joint-orthogonality tests across groups.

Table C4: Balance across assigned position of fraud

Characteristic	First	Second	Third	p-value
Female	53 %	54 %	55 %	0.526
Male	47 %	45 %	45 %	0.525
Age: 18-24	13 %	14 %	13 %	0.462
25-34	24 %	23 %	24 %	0.581
35-44	19 %	20 %	19 %	0.712
45-54	18 %	18 %	20 %	0.086
55-64	16 %	15 %	14 %	0.064
65+	10 %	10 %	10 %	0.928
Region: East Midlands	7 %	7 %	8 %	0.903
East of England	8 %	8 %	8 %	0.874
Greater London	13 %	14 %	13 %	0.310
North East	5 %	4 %	4 %	0.729
North West	12 %	12 %	11 %	0.456
Northern Ireland	3 %	2 %	2 %	0.100
Scotland	8 %	9 %	10 %	0.224
South East	15 %	14 %	15 %	0.593
South West	8 %	8 %	8 %	0.962
Wales	4 %	4 %	4 %	0.737
West Midlands	9 %	9 %	8 %	0.595
Yorkshire and the Humber	9 %	9 %	9 %	0.874
Income: £20k or less	42 %	42 %	43 %	0.774
£20-£40k	40 %	40 %	39 %	0.698
£40-£60k	12 %	11 %	12 %	0.280
£60k+	6 %	7 %	6 %	0.598
Weekly use bank app	74 %	75 %	75 %	0.385
Daily use bank app	57 %	55 %	55 %	0.267
N	3019	2976	2963	

Notes: P-values are from joint-orthogonality tests across groups.

Table C5: Treatment effects on payment behaviour

	% made a fraudulent payment		% made a legitimate payment	
	(1)	(2)	(3)	(4)
Risk-based	-0.01	0.01	0.11***	0.08***
	(0.01)	(0.02)	(0.01)	(0.02)
CTA	-0.15***	-0.12***	-0.15***	-0.14***
	(0.01)	(0.02)	(0.01)	(0.02)
Behavioural messages	-0.01	-0.04**	-0.03***	-0.06***
	(0.01)	(0.02)	(0.01)	(0.02)
Risk-based # CTA		-0.07***		-0.00
		(0.02)		(0.02)
Risk-based # Behavioural messages		0.03		0.08***
		(0.02)		(0.02)
CTA # Behavioural messages		0.02		0.01
		(0.02)		(0.02)
Risk-based # CTA # Behavioural messages		0.01		-0.04
		(0.03)		(0.03)
Constant	0.22***	0.22***	0.55***	0.57***
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	8958	8958	8958	8958
R-squared	0.045	0.048	0.057	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). Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table C6: Effects of different accuracy levels in risk-based journeys on payment behaviour

	% made a fraudulent payment (1)	% made a legitimate payment (2)
High accuracy risk-based journeys	0.00	0.03**
	(0.01)	(0.01)
Constant	0.14***	0.56***
	(0.01)	(0.01)
Observations	4490	4490
R-squared	0.000	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). High accuracy risk-based journeys is a dummy variable that takes the value of 1 when the risk-based approach did not misclassify high-risk scenarios as low-risk or vice versa, and 0 when the risk-based approach produced some false positives (i.e., classified legitimate scenarios as being risky). Standard errors in parenthesis. Stars represent p-values * $p<0.1$ ** $p<0.05$ *** $p<0.01$.

Table C7: Heterogeneity effects across demographic groups for fraudulent payments

	% made fraudulent payment										
	Age group			Income				Web banking weekly		App banking weekly	
	18 - 34	35 - 54	55+	£20k or less	£20k-£40k	£40k-£60k	£60k and over	not use	use	not use	use
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Control + behavioural	-0.01	-0.04*	-0.08***	-0.03	-0.03	-0.13***	-0.05	-0.08**	-0.03	-0.02	-0.05**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.07)	(0.03)	(0.02)	(0.02)	(0.02)
Control + CTA	-0.12***	-0.11***	-0.16***	-0.10***	-0.13***	-0.17***	-0.14**	-0.15***	-0.12***	-0.09***	-0.15***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Control + behavioural + CTA	-0.15***	-0.11***	-0.18***	-0.12***	-0.14***	-0.20***	-0.22***	-0.18***	-0.13***	-0.11***	-0.17***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based	-0.03	0.03	0.02	-0.00	0.04	-0.04	-0.03	-0.03	0.02	0.01	0.01
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.05)	(0.07)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based + behavioural	-0.01	0.02	-0.00	0.01	0.00	-0.04	0.02	-0.05	0.02	0.02	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.07)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based + CTA	-0.19***	-0.16***	-0.20***	-0.18***	-0.17***	-0.22***	-0.22***	-0.20***	-0.18***	-0.14***	-0.22***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based + behavioural + CTA	-0.15***	-0.15***	-0.19***	-0.16***	-0.15***	-0.19***	-0.23***	-0.19***	-0.15***	-0.12***	-0.19***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)
Constant	0.24***	0.20***	0.22***	0.22***	0.21***	0.25***	0.25***	0.23***	0.22***	0.18***	0.26***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.05)	(0.03)	(0.01)	(0.02)	(0.02)
Observations	3333	3390	2235	3779	3558	1058	563	2295	6663	4003	4955
R-squared	0.042	0.046	0.072	0.042	0.051	0.056	0.081	0.049	0.049	0.035	0.059

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). Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table C8: Heterogeneity effects across demographic groups for legitimate payments

	% made legitimate payment										
	Age group			Income				Web banking weekly		App banking weekly	
	18 - 34	35 - 54	55+	£20k or less	£20k-£40k	£40k-£60k	£60k and over	not use	use	not use	use
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Control + behavioural	-0.05**	-0.05*	-0.10***	-0.10***	0.01	-0.13***	-0.14**	-0.11***	-0.04**	-0.07***	-0.05**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.07)	(0.03)	(0.02)	(0.02)	(0.02)
Control + CTA	-0.16***	-0.13***	-0.15***	-0.12***	-0.12***	-0.24***	-0.27***	-0.15***	-0.14***	-0.15***	-0.14***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Control + behavioural + CTA	-0.22***	-0.19***	-0.22***	-0.19***	-0.16***	-0.29***	-0.32***	-0.22***	-0.20***	-0.22***	-0.18***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based	0.02	0.08***	0.15***	0.08***	0.13***	-0.09*	0.01	0.06*	0.08***	0.08***	0.07***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.05)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based + behavioural	0.04*	0.13***	0.14***	0.12***	0.11***	-0.02	0.10*	0.07**	0.11***	0.11***	0.09***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.05)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based + CTA	-0.08***	-0.05**	-0.08**	-0.08***	-0.03	-0.14***	-0.12*	-0.07**	-0.07***	-0.06**	-0.08***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.05)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Risk-based + behavioural + CTA	-0.09***	-0.05*	-0.11***	-0.09***	-0.02	-0.21***	-0.12*	-0.13***	-0.06***	-0.05**	-0.10***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Constant	0.66***	0.55***	0.46***	0.56***	0.53***	0.68***	0.63***	0.52***	0.58***	0.58***	0.56***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.05)	(0.02)	(0.01)	(0.02)	(0.02)
Observations	3333	3390	2235	3779	3558	1058	563	2295	6663	4003	4955
R-squared	0.049	0.059	0.098	0.065	0.055	0.059	0.120	0.061	0.061	0.071	0.052

Notes: The regressions were conducted using a Linear Probability Model (LPM). The outcomes are the share of legitimate payments made per participant (semi-continuous variable that can take the values 0, 0.5, or 1). Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table C9: Interaction effects of order of fraud and treatment elements

	% made a fraudulent payment			% made a legitimate payment		
	Fraud first	Fraud second	Fraud third	Fraud first	Fraud second	Fraud third
	(1)	(2)	(3)	(4)	(5)	(6)
Control + behavioural	-0.08***	-0.04	-0.00	-0.07**	-0.05*	-0.06**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Control + CTA	-0.15***	-0.12***	-0.10***	-0.14***	-0.12***	-0.17***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Control + behavioural + CTA	-0.20***	-0.12***	-0.11***	-0.21***	-0.15***	-0.25***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Risk-based	-0.02	0.03	0.02	0.04	0.11***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Risk-based + behavioural	-0.05*	0.04	0.02	0.05	0.14***	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Risk-based + CTA	-0.22***	-0.16***	-0.17***	-0.11***	-0.03	-0.07**
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Risk-based + behavioural + CTA	-0.18***	-0.17***	-0.13***	-0.12***	-0.05*	-0.07**
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Constant	0.27***	0.21***	0.18***	0.60***	0.54***	0.57***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	3019	2976	2963	3019	2976	2963
R-squared	0.049	0.054	0.047	0.046	0.057	0.081

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). Columns indicate whether participants were exposed to fraud in the first, second, or third scenario. Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table C10: Effects on customer satisfaction measures

	Share that would prefer to use this app over current banking app	Share that would be likely to recommend app to friends/family	Share that agree that the app felt intuitive	Share that agree that the app was easy to use	Share that agree that the app felt safe	Share that agree that the app had an unnecessary number of steps	Share that read the text and warnings presented in the app	Share that agree that it was easy to cancel payments
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Control + behavioural	0.01	0.85	0.00	0.02	0.03*	0.00	0.04***	0.03**
	(0.02)	(1.17)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Control + CTA	-0.01	0.66	0.01	0.02	0.04*	-0.01	0.03**	0.07***
	(0.02)	(1.17)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Control + behavioural + CTA	0.00	1.86	0.03	0.05***	0.04*	0.01	0.04***	0.10***
	(0.02)	(1.15)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Risk-based	0.01	1.75	0.01	0.01	0.01	0.02	0.03**	-0.04**
	(0.02)	(1.15)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Risk-based + behavioural	0.01	-0.81	0.00	-0.02	0.01	0.03*	0.02*	-0.04**
	(0.02)	(1.17)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Risk-based + CTA	0.00	-0.30	0.03	0.04**	0.03	0.01	0.03***	0.09***
	(0.02)	(1.15)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Risk-based + behavioural + CTA	0.04*	1.81	0.04**	0.03*	0.05**	0.03	0.03**	0.08***
	(0.02)	(1.16)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Constant	0.68***	57.30***	0.65***	0.77***	0.67***	0.25***	0.89***	0.83***
	(0.01)	(0.82)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	8958	8957	8958	8958	8958	8958	8958	8958
R-squared	0.001	0.001	0.001	0.002	0.001	0.001	0.002	0.025

Notes: The regressions were conducted using a Linear Probability Model (LPM). The outcomes are binary values from agree/disagree statements with 5-point Likert responses. We code as 1 if the participant states that they "Strongly Agree" or "Agree" with the statement and 0 otherwise. Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.

Table C11: Effects on time spent on scenarios

	Seconds spent on all three scenarios (1)	Average seconds spent per scenario (2)
Control + behavioural	442.89	147.63
	(912.47)	(304.16)
Control + CTA	-615.06	-205.02
	(479.34)	(159.78)
Control + behavioural + CTA	420.88	140.29
	(884.92)	(294.97)
Risk-based	-724.13	-241.38
	(471.53)	(157.18)
Risk-based + behavioural	714.23	238.08
	(830.09)	(276.70)
Risk-based + CTA	-123.67	-41.22
	(581.75)	(193.92)
Risk-based + behavioural + CTA	-577.37	-192.46
	(487.69)	(162.56)
Constant	1239.50***	413.17***
	(471.38)	(157.13)
Observations	8957	8957
R-squared	0.001	0.001

Notes: The table displays the effects on time spent on all three scenarios. Standard errors in parenthesis. Stars represent p-values * p<0.1 ** p<0.05 *** p<0.01.