Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Navigating the Rise in Non-Institutional Digital Fraud: An Experiment with Micro Enterprises in Nigeria

Shane Byrne, Daniel Putman, Michael King and Chaning Jang

TEP Working Paper No. 1224

November 2024

Trinity Economics Papers
Department of Economics

Navigating the Rise in Non-Institutional Digital Fraud: An Experiment with Micro Enterprises in Nigeria*

Shane Byrne† Daniel Putman‡ Michael King§ Chaning Jang¶

November 2024

Abstract

The high prevalence of digital financial fraud stresses businesses' ability to distinguish between real communications from digital financial service (DFS) providers and fraudulent impersonations. Besides the financial and psychological costs to businesses, fraud may erode trust in, and usage of DFS. We test two strategies for preventing non-institutional fraud: a series of anti-fraud learning interventions and a technical solution to authenticate inbound communications from a digital platform. Using a pre-registered behavioural laboratory experiment in Nigeria, we find evidence that timely educational interventions increased trust in DFS, its likely future usage, and improved knowledge about fraud four weeks post intervention. However, when we task micro business owners with evaluating the authenticity of a series of fictionalised scenarios, we do not find evidence of improvement in fraud detection, either overall, or when considering only genuine or fraudulent scenarios. Surprisingly, we find increased self-confidence in fraud detection ability, highlighting the risk of overconfidence.

JEL Codes: D18, G41, G53, O12

Keywords: Financial Behavior; Digital Finance; Fraud; Trust; Consumer Protection; Financial Inclusion; Financial Development.

*For helpful comments and valuable feedback, we thank participants of the Consumer Protection Sessions at the IPA Researcher Gathering in 2021 and 2022. Thank you to Prof. Kabir, Nasir Adam and Ismail Sanusi from ABU for their valuable planning, coordination, research assistance and use of Lab facilities. Thank you to Hafsah Jumare, Luqman Hussein, Kelechi Ogbuku, and Usman Musa from Amana Market for overall project design, recruitment and feedback on the experiment. Thank you to Ajibola Kolajo, Segun Alex Fayomi, Suleiman Amanela, Mathilde Schilling, Aya Vang, Ruth Wambua, Nathaniel Peterson, and Louis Graham from Busara for inputs to study design, qualitative and background research, survey and experimental design, experimental implementation and research assistance and partnership management. All errors and omissions are the authors' own. We thank the IPA Consumer Protection Initiative and the Bill & Melinda Gates Foundation for generous funding of this research.

†Central Bank of Ireland

‡University of Pennsylvania Center for Social Norms and Behavioral Dynamics

§Trinity College Dublin, the University of Dublin

¶Busara Center for Behavioral Economics

1 Introduction

Non-institutional fraud targeted at micro and small enterprises (MSEs) is pervasive across low- and middle-income countries (LMIC) and has risen in the wake of the COVID-19 pandemic (Tade, 2021; Kabanda et al., 2018; Fu and Mishra, 2022). Non-institutional fraud can include phishing,[1] scams to access passwords and log-ins, impersonating a formal institution, offering fake products or services and absconding with payments, and using psychological manipulation to persuade victims to part with money (Garz et al., 2021; Titus et al., 1995).[2] While digital finance services (DFS) have increased financial inclusion, it also has been linked to fraud (Erel and Liebersohn, 2022; Griffin et al., 2023). In surveys of DFS users, 56% of Kenyan respondents, 33% of Ugandan respondents, and 42% of Nigerian respondents had faced phishing scams in the months after the start of the pandemic (Blackmon et al., 2021a,b; Bird and Mazer, 2021). The presence of these scams was the most prevalent cited challenge faced in consumer engagement with digital financial services (DFS) in Kenya and Uganda, and the third most prevalent in Nigeria.

Non-institutional fraud (sometimes called third-party fraud) causes immediate and long-term damage. Immediately, it can lead to monetary loss, but also to psychological impacts including anger, difficulties with trust, feelings of violation, stress, and social embarrassment (Vohs et al., 2007; DeLiema et al., 2017). Additionally, because being defrauded is a violation of trust, it may erode trust in counterparties and institutions, a key driver of growth (Francois and Zabojnik, 2005; Gurun et al., 2018). Specifically, it can frustrate economic development through the suppressed use of digital services and under-exploitation of advantageous opportunities (Caribou Digital, 2019; Banerjee et al., 2019).[3] Likewise, when fraud operates through interpersonal networks, this might encourage the pursuit of opportunities only within closely trusted networks (Nash et al., 2013).[4]

There is limited knowledge on what mitigation strategies can be taken to reduce fraud, effectively signal authenticity and engender trust, and thereby facilitate the better realisation of the promise of digital financial services to MSEs in LMICs. Despite this, the economics of fraud are closely related those of adverse selection. Fraud takes place in settings where there is asymmetric information over the quality (and in many cases, existence) of goods, services,

[1] The fraudulent practice of sending emails purporting to be from reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers

[2] At the core, non-institutional fraud is carried out by individuals or groups who are not affiliated with a formal institution who seek to trick victims into directly sending money, or sending sensitive information that can be used to defraud the victim.

[3] Notwithstanding regulatory differences, the lower rate of digital payments in Nigeria compared to Kenya may reflect lower levels of trust in digital financial service (World Bank Group, 2019).

[4] If this limits the number of suppliers one purchases from, for example, it may lead to monopolistic competition (i.e., differentiation along the dimension of trust) and could lead to higher prices.

or financial rewards (Akerlof, 1970; Miles and Pyne, 2017).[5] While some instances can persist longer, the great majority of non-institutional fraud is designed for single interactions with unsuspecting victims (Titus et al., 1995; Herley, 2012).[6] Therefore, we test two interventions designed to address the information asymmetries that lead to fraud victimization by improving detection of fraudulent messaging. Similar interventions, particularly related to financial literacy, have been widely used.[7]

Existing efforts to combat the threat of fraud have centered variously on educational interventions which seek to arm users with the means to recognise and sidestep attempted deception (Burgoon, 2015), technological solutions to verify counterparty identities and authenticate communications (Conroy, 2017), and centralised algorithmic tools used to detect and flag anomalous patterns of behaviour (Hilal et al., 2022).

In this paper, we analyse results from a framed field experiment conducted in partnership with Amana Market, a digital market platform for agricultural trading, with a sample of 780 participants from the partner's existing network in Nigeria (Harrison and List, 2004). We test the impact of a series of learning interventions in improving the ability of small business owners to accurately discern fraudulent and genuine communications, as well as in building trust in DFS.[8,9] We also test the potential for a technical solution for the authentication of inbound communications to establish confidence and engagement (a 'unique communications code', or UCC). Our learning interventions are just-in-time, range in intensity, and are designed variously to forewarn and encourage vigilance in respect of fraud, up to arming business owners with key signs to watch out for with applied illustrative examples. Our most intensive intervention (Treatment 3) lasts a total of 25 minutes. As fraud victimization is notoriously difficult to study in the field, we take advantage of the behavioral lab setup to study fraud detection ability.[10] As part of an experimental task, participants are

[5]Some fraud and scams are related to adverse selection in credit markets since they depend on promises of the delivery of goods in the future (Stiglitz and Weiss, 1981). This includes schemes like advanced fee fraud.

[6]Longer term fraud in competitive markets requires either the joint diagnosis and provision of services (e.g., auto-mechanic work) or other types of credence goods (e.g., antivirus software) (Darby and Karni, 1973; Stone-Gross et al., 2013).

[7]A vast array of generic, general audience public information campaigns have developed around the world to encourage awareness in relation to digital fraud. These include efforts from the US Federal Trade Commission (`https://consumer.ftc.gov/articles/how-avoid-scam`, `https://consumer.ftc.gov/features/fotonovelas`), UK Finance (`https://www.takefive-stopfraud.org.uk/toolkit/`, `https://quiz.takefive-stopfraud.org.uk/`), Banco de Portugal (`https://clientebancario.bportugal.pt/en/material/5-tips-staying-safer-online-toptip`) among many many others.

[8]This experiment was conducted in partnership with Amana Market, a digital platform in Nigeria that offers access to market information and financial service to MSEs.

[9]The experiment was pre-specified (Byrne et al., 2022) and pre-registered with the AEA RCT Registry (RCT ID: AEARCTR-0009470).

[10]Since fraudsters seek to deceive, the may induce survey noise which extends beyond recall bias. Further-

required to evaluate 20 fictionalised communications scenarios for their authenticity, and report their subjective feeling of confidence in their judgements. To complement our core experimental evaluation, we use data from baseline and endline surveys to explore treatment effect heterogeneity across relevant demographic, behavioural, and experiential factors. We conduct a follow-up knowledge retention quiz four weeks after participants have left the lab.

We do not find evidence that these learning interventions significantly improve discriminant ability between genuine and fraudulent communications, although our coefficients on these interventions are positive and trending upwards as we deepen the educational intervention, suggesting that it could be large standard errors or low power preventing the detection of statistically significant impact. Nor do we find evidence that our UCC authentication solution acts as a sufficiently strong signal to increase user engagement with good faith customer outreach. We observe significant increases in the confidence that treated users report in their judgements, notwithstanding the absence of any corresponding improvement in actual underlying accuracy. In this, we highlight the potential for false confidence effects from ineffectual learning interventions, which may engender the subjective feeling of competence, and unintentionally increase susceptibility to fraud victimization through complacency.[11] We do, however, find positive treatment effects on other associated outcomes. We find a significant impact from treatment on in trust in DFS, potentially reflecting a heightened confidence on the part of treated participants in their ability to discern fraud, successfully navigate the digital financial landscape, and as such engage with confidence with legitimate digital financial counterparties. In addition, we find evidence of increased likelihood of future use of banks and mobile banking, as well as improved knowledge regarding the signs of fraud.

Our paper's contribution to the literature is fourfold. First, we contribute the literature on anti-fraud learning interventions, financial literacy interventions, and fraud prevention strategies more broadly. A preponderance of existing literature on anti-fraud interventions is focused on high-income country contexts, or with one-size fits all interventions for universal consumption (Fernandes et al., 2014; Kaiser and Menkhoff, 2017). In contrast, our paper undertakes an experimental intervention of anti-fraud learning among business owners in a LMIC, providing evidence for a low-income segment within an LMIC. Gathering evidence among those who are low income in LMIC contexts is particularly important given that while effects are stronger as country income falls, they tend to attenuate among those with

more, survey reports of fraud losses may rely on a minority of upper tail responses (Florêncio and Herley, 2013). Finally, complaints data may be biased by transaction costs, capacity, or other behavioral factors such as self-efficacy (Ba, 2018; Innovations for Poverty Action, 2021b; Raval, 2020).

[11]While some theoretical work has dealt with the unintended consequences of fraud deterrence, our research places this firmly within a the more common everyday context of financial literacy interventions as opposed to detecting and arresting fraudsters (Miles and Pyne, 2017).

low socio-economic status within high income countries. Additionally, our interventions are specifically adjusted to resonate within the information-environment and digital financial landscape faced by small business owners in a LMIC. In contrast to results from the broader financial literacy literature, where stronger effects are found in lower income country contexts, we do not find a strong effect in our population of small business owners in Nigeria. This finding is consistent with results from Kubilay et al. (2023), co-current research which does not find an effect of anti-fraud education on susceptibility in an online sample from Kenya. In contrast to Kubilay et al. (2023), we document confidence, trust, and usage of DFS in addition to susceptibility.

Second, we contribute to the literature on financial inclusion in LMICs. While not a panacea for poverty, financial technology adoption remains important. Microfinance may allow firm growth for experienced entrepreneurs and has proven important for aggregate economic activity (Banerjee et al., 2021; Meager, 2019; Breza and Kinnan, 2021). Index insurance products may allow for profitable risk taking in agriculture (Elbers et al., 2007; Karlan et al., 2014). DFS products like mobile money and digital credit have proved important for resilience to economic shocks (Jack and Suri, 2014; Riley, 2018; Suri et al., 2021). Our intervention increases self-efficacy in detecting fraud, and in turn, leads to greater self reported trust and future usage of financial technologies. This suggests that digital fraud serves as a constraint to trust and eventual adoption of financial services. Notably, we find fairly large effects in our self-reported measures of trust and future usage: 10% and 7% increases in the likelihood of using banks and mobile banking in the future, respectively; and a 0.2 SD increase in our index of trust in DFS. This suggests such interventions may be useful as a strategy to boost financial inclusion.

Third, our paper contributes to our understanding of the subtle dynamics of the role of confidence in the ability to detect deception. Higher confidence tends to be strongly associated with improved task performance, an effect often attributed to increased self-efficacy Bandura (1977); Bandura et al. (1999); Stankov and Crawford (1996); Woodman and Hardy (2003). However, in some cases such confidence may give rise to complacency, risk-taking, and reduced effort towards a task which can negatively impact upon performance, or conversely the introduction of doubt can result in increased effort and focus (Vancouver et al., 2001; Vancouver and Kendall, 2006; Woodman et al., 2010). In our experiment, we find that increases in confidence regarding task performance depart from actual performance. Despite the gains in trust that we document, this untethering could mean that MSEs become overconfident. This may be cause for concern: overconfidence has been shown to drive drive poor decision-making in investment markets in addition to reducing inoculation against deceptive messages (Statman et al., 2006; Ben-David et al., 2013; Lyons et al., 2021; Serra-

Garcia and Gneezy, 2021; Walters and Fernbach, 2021).

Fourth, considering the trade-off between over-confidence in fraud detection and trust in digital financial services contribute to the literature on the unintended consequences of fraud deterrence. One example of these unintended consequences is from Miles and Pyne (2017), which uses applied theory to identify a way in which arrest of fraudsters might reduce the number of low ability fraudsters, making it so that potential victims are not inoculated by feeble attempts. Kubilay et al. (2023) finds that while DFS users do not improve in detecting fraud, they do become more skeptical of both legitimate and fraudulent communications. Ultimately, the welfare effects of educational interventions will depend on the prevalence of fraud in the economy and its costs relative to the benefits of DFS.

Our failure to achieve meaningful treatment effects from learning interventions in respect of our primary detection outcome is in keeping with the relatively underwhelming pattern of results found in the literature on anti-fraud learning interventions, where no or modest effects have been frequently observed (Fernandes et al., 2014). It is, nonetheless, surprising in view of the direct nature of the instruction, the contextually-adjusted and engaging nature of the interventions, and the immediacy of the experimental task that followed. Our results speak to the severity of the challenge that small business owners are likely to face in successfully navigating the noisy landscape of competing communications, and cast doubt on the utility of relatively light-touch, quick-fix learning interventions as meaningful antidotes, even when delivered in a timely fashion. We cannot, however, exclude the possibility that true effects from our learning intervention may fall below our minimum detectable effect.[12] Equipping LMIC small business owners to successfully navigate the contemporary torrent of digital fraud to safely exploit the promise of digital financial markets is evidently a steep challenge. Our results also highlight the risk of false confidence effects from ineffectual learning interventions which may engender the feeling but not the reality of heightened competence. As such, they can be offered as a cautionary lesson for policy in this domain, which recommends introspection with respect to the type of interventions tested here, and their content and intensity. Our results also highlight the critical importance of rigorous pre-testing of planned interventions in this domain to establish what works, and just as importantly, what doesn't.

The paper proceeds as follows: Section 2 provides an overview of common strategies to combat non-institutional fraud, Section 3 describes the experimental design and empirical strategy, Section 4 reports our empirical results, and Section 5 concludes.

[12] Albeit, the economic significance of treatment effects in this range would be limited.

2 Strategies for Reducing Non-Institutional Fraud

2.1 Financial Education

2.1.1 The impact of financial education on knowledge and behavior

One of the most common approaches to reducing non-institutional fraud is the use of financial education interventions, which are also widely studied outside of the context of of financial fraud. Meta analyses have shown a positive effect on financial knowledge and behaviors, particularly when focusing on experimental evidence. In particular, Kaiser and Menkhoff (2017) and Kaiser et al. (2021) find that there are positive impacts on financial behavior from financial education initiatives, estimating effect sizes of 0.08 SD and 0.1 SD respectively.[13] In observational studies, evidence is more mixed with some behaviors unchanged or unexplained (Miller et al., 2015; Fernandes et al., 2014).[14] The effects of these programs tend to vary by context: While Fernandes et al. (2014) found weaker effects from initiatives in low-income samples within country, countries with higher average income (and education) have weaker effects as well (Kaiser and Menkhoff, 2017).[15]

Some aspects of program design matter. In particular, there seem to be returns to intensity, with more hours of education resulting in larger effects (Fernandes et al., 2014; Kaiser and Menkhoff, 2017).[16] Likewise, the timing of interventions matters as well as the effects tend to decay over time (Fernandes et al., 2014; Kaiser et al., 2021).[17] Fernandes et al. (2014) envisage a role for 'just-in-time' financial education tied to the specific behaviours it intends to impact while Kaiser and Menkhoff (2017) point to providing financial education at a 'teachable moment' (i.e., when teaching is directly linked to decisions of immediate relevance to the target group). There is little evidence for other aspects of the interventions mattering (Miller et al., 2015; Kaiser and Menkhoff, 2017).[18]

[13]Kaiser and Menkhoff (2017) aggregates across 126 studies and Kaiser et al. (2021) aggregates across 76 studies. Both put higher weight on experimental evidence, with Kaiser et al. (2021) using only experimental evidence. Interestingly, the authors note that this is similar in magnitude to effect sizes reported in meta-analyses of behaviour change interventions in other domains such as health or energy conservation.

[14]Fernandes et al. (2014), aggregates 168 cases and Miller et al. (2015) aggregates across 188 cases.

[15]This may be attributable to diminishing marginal returns to additional financial education.

[16]Miller et al. (2015) finds only mixed evidence on this point. In particular, they find that intensity was weakly significant in some specifications of the model it was not significant in the others.

[17]More specifically, Fernandes et al. (2014) observe equal effects for 6 hours of intervention at no delay and 18 hours of intervention at 10 months of delay, and equal effects of 1 hour of instruction at no delay and 12 hours at 10 months of delay. Even large interventions with many hours of instruction have negligible effects on behaviour 20 months or more from the time of intervention. Kaiser et al. (2021) find a less rapid decay in treatment effects, though still little support for the long-run sustainability of effects.

[18]These include: the age and gender of participants, the delivery channel, duration exposed to the treatment, whether the intervention was staged at school, in the community, or in the workplace.

2.1.2 Educational interventions to curb susceptibility to financial fraud

The high social cost associated with fraud, and the difficulty of deception detection has led to significant attention to for fraud-specific educational interventions as well as others targeted at detecting deception (Burgoon, 2015). Results are mixed among this diverse array of educational interventions.[19] In some studies, interventions reduced susceptibility to fraud or deception (Biros et al., 2002; Anderson, 2003; Scheibe et al., 2014; Xiao and Benbasat, 2015; Burke et al., 2020). Other studies have produced null effects (George et al., 2004; Grazioli and Wang, 2001). Notably, [co-current study] does not find an overall impact on fraud detection in Kenya. However, simply focusing on overall accuracy—success in identifying fraud—may mask varying success in identifying fraud vs non-fraud messages. In particular if interventions increase skepticism, they may increase 'true positives'—correctly identification of fraudulent communications—at the cost of 'true negatives'—incorrectly identifying non-fraudulent communications as fraudulent (Burgoon et al., 1994; Xiao and Benbasat, 2015). As in the broader financial education literature, there is evidence of decay of effects. In particular, effects have been documented at five weeks (Anderson, 2003; Scheibe et al., 2014) and even out to six months (Burke et al., 2020).

2.2 User-centered digital tools to combat fraud

Quite apart from educational initiatives designed to arm users with the know-how to discriminate effectively between genuine and fraudulent activity, is a vast array of digital tools have been devised with the objective of detecting and preventing fraud in digital commerce. Rather than relying on training a market of discerning users with an adequate filter to separate fraudulent from genuine communications, digital tools offer the potential for variously 'smart' and automated solutions which efficiently perform that function on consumers' behalf. These solutions can be broadly classified into two camps: detection, or the ability to identify suspicious patterns indicative of fraudulent activity, and authentication, or establishing the provenance of transactions or communications. Here we focus on user centered authentication tools.[20]

Authentication tools seeking to establish that a product user is who they claim to be are traditionally achieved by validating something the user has (possession), something they are (inherence), or something they know (knowledge) (Velásquez et al., 2018). These include

[19]These include traditional training sessions, warning messages, consumer advice, and decision-making heuristics aimed at improving participants' performance in fraud detection in various digital settings.

[20]Beyond these, and of less immediate relevance to our purpose, is a set of other institutional tools used to detect financial fraud. Hilal et al. (2022) provides an instructive survey of machine-learning algorithmic anomaly detection methods in the field of financial fraud, where the majority of applications are found in insurance and credit card markets.

personal log-in credentials (i.e., user name and password), two-factor authentication, physical biometrics, and know your customer (KYC) protocols. The these tools are designed to balance the need to to stay ahead of fraudsters whose *modus operandi* evolves to erode the integrity of the defensive protocols against the need to provide frictionless user experience (Herley, 2009). Each comes with advantages and weaknesses, and can often be undermined by data breaches, creative social engineering on the part of fraudsters, or poor practices on the part of users. Frequently they are used in various overlapping layers and combinations as a means of reinforcing the reliability of gate-keeping. Increasingly, vendors are seeking to flexibly match the level of friction imposed by authentication layers to the risk of the underlying transaction, and to the preferences of the user (Conroy, 2017).

3 Experimental design

3.1 Context, sample, and session protocol

We investigate the effects of educational interventions on fraud detection, confidence, and trust, using a laboratory experiment. To conduct the experiment we worked with Amana Market, the Busara Center for Behavioral Economics and the experimental lab at Ahmadu Bello University (ABU) in Zaria, Nigeria. Zaria has a population 700,000 people and is located Kaduna State in Northern Nigeria. Amana Market recruited participants from communities in Zaria using their existing agent network. Participants were either Amana Market users or similar in profile to Amana Market users, thus predominantly those within the agricultural value chain. The profiles of the participants (name, phone number, gender, etc.) were uploaded into a database and then randomly assigned to a given lab session. Those assignments were then sent to the Amana Market agent who would bring participants to the experiment at the designated date and time.

A total of 780 participants across 52 lab sessions participated in the main experiment, which ran from July 25th - August 26th, 2022. Three pilot sessions were held on 21st July, 2022, which included 45 participants across the three treatment arms. A total of 15 participants were invited to the lab for each session. In the waiting room, participants verified their identity with a staff member to ensure they were participating in the correct session, and were randomly assigned a seat number in the lab. Upon entering the lab, informed consent was obtained from each participant. During the experiment, participants first underwent a baseline survey, followed by an educational intervention or control condition, followed by completing fraud detection experimental task and an endline survey. Upon finishing the endline survey, a UCC was created for participants to be used for the follow up process

and survey. Finally, participants are paid a total of NGN 4,500 in cash as a transport reimbursement and attendance incentive before leaving the lab.

After each participant visited the lab, we conducted an SMS and phone survey. Three weeks after the lab session, we sent each participant an SMS in which they are requested to respond confirming their month and year of birth. The SMS was randomly assigned to contain or not contain the UCC that was assigned to them at the end of the experiment. A week after we sent the SMS (and four weeks after the lab session), we conducted a follow up phone survey where we asked them about their experience with the SMS that was sent to them, and questions about key signs of fraud.

3.2 Experimental interventions

3.2.1 Educational interventions

All participants are randomised into either a control group or one of three treatment groups. Each of the three treatment groups receives a variation of a learning intervention aiming at helping participants distinguish between genuine and fraudulent communication (either a general warning treatment, or one of two targeted educational interventions). These simple interventions are meant to be brief and replicate common approaches used in anti-fraud campaigns and training. The control group initially receives no additional warning about fraud, while the treatment arms receive some warning or education. The four experimental arms are presented in Table 1.[21]

3.2.2 Unique Customer Code (UCC) intervention

At the end of the experiment, all participants were assigned a UCC. Participants are randomly allocated into one of two equally weighted groups: the non-personalised UCC group (these subjects are assigned a randomly generated 5-digit UCC), and the personalised UCC group (these subjects are instructed to choose their own 5-digit UCC). The lab staff explain that this code will be used to verify the authenticity of future communications with

[21]It could be argued that our control group is inherently primed to think about fraud by virtue of being confronted with the experimental task where the evaluation of authenticity and fraudulence is the clear objective, and as such that the control contains an element of implicit treatment which may serve to mute the additional treatment effect from the subsequent learning interventions. Such implicit treatment is inescapable in our setting, and will be consistent across treatment arms, meaning that the additional benefit of the learning interventions under evaluation will still be identified. However, the possibility of such a placebo effect will mean that our estimated treatment effects cannot be treated as an ecologically valid estimate of a treatment effect that might be observed in the field, where the control benchmark condition involves no such implicit task-based priming.

Table 1: Anti-Fraud Campaign Interventions

Control (C)	Control group receive the lab manager's session introduction and undergo the consenting process, but receive no additional warning or educational information related to fraud.
Treatment 1 (T1)	On top of the lab manager's session introduction, and the consenting process, T1 subjects receive on-screen general warning messages stating, "Digital fraud represents a threat to small businesses in Nigeria. Fraudsters may contact you pretending to represent legitimate businesses or agencies, in an effort to take your information or your money. Be on the lookout for signs of potential fraudsters in the communications you receive â over the phone, by email, or in person".
Treatment 2 (T2)	On top of the lab manager's session introduction, and the consenting process, T2 subjects receive an on-screen written list of 7 key signs of potential fraud which is narrated in an audio file (key signs are detailed in Table 26. This information is prefaced by a general warning message (see Treatment 1). To aid recall, subjects are prompted to write down the key signs upon completion, before replaying the 7 key signs and filling in the gaps in their answer sheets. Subjects' notes are collected before the remainder of the lab session.
Treatment 3 (T3)	On top of the lab manager's session introduction, and the consenting process, T3 subjects receive an on-screen written list of 7 key signs of potential fraud, complemented with applied illustrative examples which is narrated in an audio file. This information is prefaced by a general warning message (see Treatment 1). To aid recall, subjects are prompted to write down the key signs upon completion, before replaying the 7 key signs and filling in the gaps in their answer sheets. The subject's notes were collected before the remainder of the lab session.

Notes: The table describes the four groups into which subjects are randomly assigned.

participants. All codes are recorded centrally, and sent to subjects by SMS to keep as a record.

3.2.3 Randomisation and groups

Participants are randomly assigned at three independent stages during the lifetime of the experiment. First, participants are randomly assigned into one of three educational interventions or the control group. Second, participants are assigned to either receive an automatically generated 5-digit UCC, or to personalise their own 5-digit code. Finally, participants are assigned to either receive an SMS which has their UCC embedded, or which does not contain their UCC.

3.3 Experimental task, survey, and follow-up quiz

The primary outcome measure is performance on a task where participants attempt to identify fraudulent or legitimate communications from a fictitious sender. Participants are

exposed to 20 fictitious scenarios, half fraudulent and half genuine. All participants are exposed to the same 20 scenarios, with the aim of discerning genuine from fraudulent communications. The scenarios are shown to each participant in a random order. A description of scenarios presented is given in Table 27, along with illustrative examples in Figures 36 and 37. We measure both their raw performance (whether they identified the scenario correctly as fraudulent or genuine) as well as their confidence level in their answer.

We survey the participants both before and after the experimental session (i.e., treatment and experimental task). The baseline survey collects demographic information, attitudes and experience with digital financial services, previous experience with fraud, trust in DFS, likelihood to use to DFS. An endline survey again asks participants for their trust and willingness to engage in digital financial services.

We additionally investigate the degree to which participant recall of the key signs of fraud varies across treatment arms in a knowledge retention follow-up quiz administered over the phone four weeks following the completion of the main experimental task in the lab. Participants are presented with three multiple-choice questions in which they are asked to correctly identify which item represents a key sign of fraud.

3.4 Empirical strategy

To estimate the causal effect of treatments on participant ability to distinguish between fraudulent and genuine communications, we perform the following empirical specification:

$$Y_i = \alpha + \beta_1 \mathrm{T1}_i + \beta_2 \mathrm{T2}_i + \beta_3 \mathrm{T3}_i + \epsilon_i \tag{1}$$

where we define Y_i to be one of the outcome variables described in table 1 or table 2 for participant i, $\mathrm{T1}_i$, $\mathrm{T2}_i$, and $\mathrm{T3}_i$ denote treatment arms described in Table 1. β_1, β_2, and β_3 estimate the corresponding treatment effects. For each outcome, we test the three following three hypotheses:[22]

- Hypothesis 1.1: $H_0 : \beta_1 \leq 0$. Providing MSEs with a general warning message about fraud alone (with no further educational intervention) improves their ability to distinguish between genuine and fraudulent communications (T1 vs. C).
- Hypothesis 1.2: $H_0 : \beta_2 \leq \beta_1$. Providing MSEs with seven key warning signs for potential fraud in a simple format (written/audio) improves their ability to distinguish

[22]We perform the following hypothesis tests after running the regression using accuracy (i.e., accurate identification of fraud scenarios) as our primary outcome. We additionally partition the accuracy outcome to separately analyse accuracy in respect of genuine scenarios (true positives) and accuracy in respect of fraudulent scenarios (true negatives). The other hypotheses relating to confidence, trust of DFS, and likelihood of using DFS (outlined in Table 12) are tested in the same manner.

between genuine and fraudulent communications, still further than can be achieved by a general warning message alone (T2 vs. T1).

- Hypothesis 1.3: $H_0 : \beta_3 \leq \beta_2$. Illustrating applied examples of fraudulent communications in a simple format (written/audio) improves MSEsâ ability to distinguish between genuine and fraudulent communications, still further than can be achieved with simple warning signs alone (T3 vs. T2).

We also test if any of the treatments improve the ability to distinguish between genuine and fraudulent communications by running a pooled specification:

$$Y_i = \alpha + \beta_1 \mathrm{T}_i + \epsilon_i \tag{2}$$

where T_i indicates that participant i receives any of the three treatments. This specification is used to test $H_0 : \beta \leq 0$.

3.4.1 Heterogeneous treatment effects

The individuals targeted by scammers exhibit significant heterogeneity in ability to detect fraud (e.g., based on sophistication, skepticism, or experience) (Vohs et al., 2007; Holtfreter et al., 2010; Engels et al., 2020). Therefore, we explore how treatment effects may vary along important experiential, attitudinal, and demographic dimensions, which are elicited as part of the baseline survey which precedes the experimental task. For most of these outcomes, a standardized index is computed then split into types by those who are above or below average according to that index. In the cases of ICT experience, DFS experience, self-control, risk appetite, and generalised trust and skepticism it makes the most sense to compute these indices using Principal Components Analysis and taking the first component of that index, but in other cases we use a context-specific index (e.g. fraud experience).

We additionally investigate the extent to which relevant socio-demographic factors interact with the learning interventions: specifically sector of employment, age, education, and gender.

To estimate these heterogeneous treatment effects, we perform the following specification:

$$Y_i = \alpha + \beta_1 \mathrm{T1}_i + \beta_2 \mathrm{T2}_i + \beta_3 \mathrm{T3}_{ii} + \delta_1 \mathrm{T1}_i \times \mathrm{M}_i + \delta_2 \mathrm{T2}_i \times \mathrm{M}_i + \delta_3 \mathrm{T3}_i \times \mathrm{M}_i + \epsilon_i \tag{3}$$

Where M_i is the moderating variable, and δ_1, δ_2, and δ_3 represent the change in the slope of the corresponding simple effects captured by β_1, β_2, and β_3. We test the null hypothesis that the change in the slope is equal to zero in each case, i.e. $H_0 : \delta_1, \delta_2, \delta_3 = 0$.

3.4.2 UCC follow-up exercise

We evaluate the potential for the UCC to act a signal of authenticity in user communications and to elicit engagement. First, we assess how the likelihood of user engagement through SMS response is impacted by the presence of the UCC embedded in the outreach. We estimate the specification:

$$Y_i = \alpha + \eta \ \text{UCC}_i + \epsilon_i \tag{4}$$

where Y_i is a variable indicating if the participant responded and UCC_i indicates that the UCC was included in the communication. We test the null hypothesis that the coefficient is less than or equal to zero, i.e. $H_0 : \eta \leq 0$. Additionally, we will explore whether having personalised the UCC at the close of the lab session, as distinct from having one automatically assigned, strengthens the signal of authentication and further elicits user engagement. We estimate the specification:

$$Y_i = \alpha_1 + \eta_1 \text{UCC}_i + \eta_2 \text{UCC}_i \times \text{Personalised}_i + \epsilon_i \tag{5}$$

We test the null hypothesis that the coefficient is less than or equal to zero, i.e. $H_0 : \eta_2 \leq 0$.

3.4.3 Learning by doing

In addition to considering heterogeneity in learning effects by fraud experience, we explore learning by doing within the experiment, to assess whether performance improves and confidence grows over the course of the experimental task. To evaluate this, we test whether those scenarios that appeared later in the order were more often correctly identified by participants. In the same manner we test whether scenarios appearing later in the order were judged with a higher degree of confidence.

$$Y_{is} = \alpha + \theta \text{Order}_{is} + \epsilon_i \tag{6}$$

where Order_{is} is the order scenario s was presented to participant i, and Y_{is} represents the accuracy and confidence in scenario judgements. Again, we test the null hypothesis that the coefficient is less than or equal to zero, that is $H_0 : \theta \leq 0$ to test for learning in both cases.

3.4.4 Experimental balance

Table 2 reports descriptive statistics for key demographic, experiential, and attitudinal characteristics across treatment cells. In order to attribute any observed difference in specified

outcomes to the impact of the interventions under evaluation, it is important that randomisation was performed effectively, with the result that treatment groups are well-balanced in key covariates at the outset.

Following McKenzie (2015), Table 13 shows a pairwise regression of treatment status (control vs. each of our treatment groups) on the same vector of covariates included in Table 2, which may be correlated with our outcome variable of interest, to ascertain whether these factors differ systematically and help to predict treatment status. While we find a high degree of statistical balance in most cases, we do observe some evidence of significant imbalance, most notably in gender. Following Mutz et al. (2019), and to adjust our estimation for potentially confounding influence, and to obtain more precise treatment effect estimates, we include as a vector of controls in our estimation of treatment effects all those pre-specified prognostic variables of interest listed in Table 11. As an alternative approach to the selection of relevant control variables, we use partialling out lasso linear regression, which selects relevant control variables for inclusion in the estimation regression, and find our estimation is unchanged.

4 Results

4.1 Effects of fraud education interventions

4.1.1 Main effects

In this section, we outline the impact of our experimental fraud education interventions on recipients' fraud detection performance, confidence in their performance, trust in DFS providers and their likelihood to use these providers in the future, when compared against the control group.

We first show in Table 3 that no intervention succeeds in significantly impacting upon the overall level of accuracy across fraudulent and genuine scenarios. Separating performance by true positives and true negatives, we find no specific impact on true negatives, but that Treatment 3 yields a significant negative impact on true positive performance of approximately 8%. This may reflect a heightened level of skepticism engendered by exposure to treatment, with the result that genuine scenarios are more likely to be rejected as fraudulent (that is, a higher rate of 'false alarms'). Table 3 also reports tests of our pre-specified hypotheses that each successive treatment arm delivers incremental added value compared to the preceding arm in the sequence (i.e. Treatment 1 dominates Control, Treatment 2 dominates Treatment 1, and Treatment 3 dominates Treatment 2). In each case, we fail to reject the null hypothesis of no incremental benefit. We additionally investigate whether

Table 2: Descriptive statistics across treatment cells

VARIABLES	(1) Control	(2) Treatment 1	(3) Treatment 2	(4) Treatment 3
Age (Years)	26.78	26.74	26.40	28.09
	(7.11)	(6.73)	(7.19)	(7.35)
Female (%)	0.65	0.41	0.43	0.45
	(0.47)	(0.49)	(0.50)	(0.50)
Third level education (%)	0.57	0.57	0.53	0.51
	(0.50)	(0.50)	(0.50)	(0.50)
Married (%)	0.50	0.51	0.47	0.53
	(0.50)	(0.50)	(0.50)	(0.50)
Agricultural employment (%)	0.31	0.38	0.41	0.37
	(0.46)	(0.49)	(0.49)	(0.49)
Contacted by scammer (%)	0.64	0.66	0.73	0.74
	(0.48)	(0.48)	(0.44)	(0.44)
Access to smartphone (%)	0.94	0.90	0.97	0.93
	(0.24)	(0.30)	(0.17)	(0.25)
Business owner (%)	0.90	0.89	0.89	0.87
	(0.30)	(0.32)	(0.32)	(0.34)
Has formal financial account (%)	0.85	0.83	0.88	0.83
	(0.36)	(0.38)	(0.32)	(0.38)
Used online platforms (%)	0.34	0.33	0.39	0.40
	(0.48)	(0.47)	(0.49)	(0.49)
Trusting (%)	0.48	0.58	0.53	0.53
	(0.50)	(0.49)	(0.50)	(0.50)
Risk averse (%)	0.35	0.25	0.32	0.27
	(0.48)	(0.43)	(0.47)	(0.47)
Observations	195	195	195	195

Notes: Table reports means and standard deviations in parentheses of mortgage borrower characteristics in each treatment and control group.

treatment effects are in evidence when treatment arms are pooled (all treatments together, or just the more intensive treatment arms 2 and 3) in Table 24, and still find no effect.[23,24]

It is important to note that, all estimated coefficients reported in Table 3 are below our estimated ex-post minimum detectable effect in each primary outcome, on the basis of observed outcomes in the control condition (see Section A.3). As such, we are not sufficiently powered to estimate with confidence true treatment effects that fall in the region of coefficient magnitudes reported here. However, we do not regard effects that fall so far below

[23]While we do not find consistent significant evidence of treatment effects on accuracy, Figure 1 depicts how the directional pattern of estimated effects points to a dis-improvement in the rate of true positives (i.e. a higher likelihood of false alarms), and an improvement in the rate of true negatives (i.e. a higher percentage of hits) in proportion with the intensity of the treatment administered, consistent with an overall increase in skepticism towards inbound communications.

[24]In Table 14, we re-estimate results depicted in Table 3, but using partialling out lasso linear regression, and find consistent treatment effect estimates.

Table 3: Overall effect

	(1) Overall	(2) True positive	(3) True negative
Treatment 1	-0.010	-0.021	0.001
	(0.013)	(0.026)	(0.025)
Treatment 2	-0.001	-0.026	0.023
	(0.012)	(0.026)	(0.023)
Treatment 3	-0.009	-0.049*	0.032
	(0.013)	(0.026)	(0.023)
Constant	0.654***	0.591***	0.718***
	(0.016)	(0.033)	(0.030)
Observations	780	780	780
R-squared	0.085	0.019	0.062
p-value (T1≤C)	0.791	0.794	0.483
p-value (T2≤T1)	0.237	0.563	0.162
p-value (T3≤T2)	0.725	0.810	0.340

Notes: Table reports results from two-sided test for treatment effects on overall accuracy, true positives, and true negatives. Also reported are one-sided tests of pre-specified hypotheses for incremental positive treatment effects from each treatment arm compared against the preceding arm in the sequence. Regression includes vector of controls listed in Table 11. *** p<0.01, ** p<0.05, * p<0.1

our minimum detectable effects as being economically meaningful for the purposes of this experiment.

We observe in Table 4 that notwithstanding the absence of any positive impact on overall detection ability, we do find consistent positive impacts from Treatments 2 and 3 on the level of confidence that participants report in their judgements over the presented scenarios overall, and both in respect of genuine and fraudulent scenarios of 3-5%. These effects are modest, but illustrate the troubling potential for educational interventions to engender a subjective feeling of confidence with respect to a given task, without necessarily delivering any actual improvement in underlying ability and performance. Again, we additionally test for incremental impacts from each treatment arm compared against the preceding arm in the sequence. For each outcome, we fail to reject the null hypothesis of no incremental benefit from Treatment 1 compared against the control. We do, however, for each outcome reject the null hypothesis of no incremental benefit associated with Treatment 2 when set

Table 4: Confidence effect

	(1) Overall	(2) True positive	(3) True negative
Treatment 1	0.006	-0.010	0.017
	(0.115)	(0.116)	(0.120)
Treatment 2	0.207*	0.200*	0.214*
	(0.112)	(0.115)	(0.114)
Treatment 3	0.256**	0.207*	0.283**
	(0.115)	(0.118)	(0.118)
Constant	5.901***	5.858***	5.935***
	(0.143)	(0.147)	(0.148)
Observations	780	780	780
R-squared	0.164	0.152	0.159
p-value (T1≤C)	0.480	0.533	0.443
p-value (T2≤T1)	0.0265	0.0246	0.0325
p-value (T3≤T2)	0.316	0.476	0.248

Notes: Table reports results from two-sided test for treatment effects on overall accuracy, true positives, and true negatives. Also reported are one-sided tests of pre-specified hypotheses for incremental positive treatment effects from each treatment arm compared against the preceding arm in the sequence. Regression includes vector of controls listed in Table 11. *** p<0.01, ** p<0.05, * p<0.1

against Treatment 1, indicating that providing participants with warning signs for potential fraud does impact upon confidence, beyond what can be achieved by a general warning message alone. Finally, we fail to reject the null hypothesis of no further incremental benefit associated with Treatment 3 when set against Treatment 2.[25,26]

In Table 5 we evaluate the impact of our interventions on an inverse-correlation weighted matrix of reported trust in DFS.[27] In respect of Treatment 2, we find a significant and positive effect, with an approximately 0.2 standard deviation increase in the standardised inverse-correlation weighted index of trust, when compared against the control group. This finding is shown to be robust in a specification which additionally adjusts for the trust

[25]We additionally reported aggregated treatment effects when treatment arms are pooled (all treatments together, and just the more intensive treatment arms 2 and 3) in Table 25.

[26]In Table 15, we re-estimate results depicted in Table 4, but using partialling out lasso linear regression, and find consistent treatment effect estimates.

[27]See Table 28 for complete definition of all variables used.

Table 5: Effect on trust (ICW trust index)

	(1)	(2)
Treatment 1	0.135	0.055
	(0.097)	(0.076)
Treatment 2	0.211**	0.171**
	(0.095)	(0.081)
Treatment 3	0.166	0.084
	(0.102)	(0.083)
Baseline trust index		0.549***
		(0.038)
Constant	0.062	0.019
	(0.118)	(0.096)
Observations	780	780
R-squared	0.182	0.432
p-value (T1≤C)	0.0819	0.233
p-value (T2≤T1)	0.192	0.0615
p-value (T3≤T2)	0.686	0.858

Notes: Table reports results from two-sided test for treatment effects on a standardised inverse-correlation weighted index of trust in DFS. Also reported are one-sided tests of pre-specified hypotheses for incremental positive treatment effects from each treatment arm compared against the preceding arm in the sequence. Regression includes vector of controls listed in Table 11. *** p<0.01, ** p<0.05, * p<0.1

index measured ex-ante as part of the baseline survey. This suggests that once armed with experience of the anti-fraud educational interventions, participants are more likely to trust DFS providers. This may reflect a heightened confidence on the part of treated participants in their ability to discern fraud, successfully navigate the digital financial landscape, and as such engage with confidence with legitimate digital financial counterparties.

To complement our analysis of intervention impacts on endline trust, we additionally examine the impact of treatment on the likelihood of future use of various specific financial service entities in Table 6. We find evidence for a heightened likelihood of the future use of banks and mobile banking of approximately 10% and 7% respectively, with no such significant effect observed in respect of mobile money operators, online platforms, or agents. For each outcome, we fail to reject the null hypothesis of no incremental benefit from Treatment 1 compared against the control. We do, however, in respect of banks and mobile banking reject the null hypothesis of no incremental benefit associated with Treatment 2 when set

Table 6: Effect on likelihood of future use

	(1) Banks	(2) Banks	(3) Mobile banking	(4) Mobile banking	(5) Mobile money	(6) Mobile money	(7) Online platforms	(8) Online platforms	(9) Agents	(10) Agents
Treatment 1	-0.078	-0.165	-0.182	-0.258	-0.125	-0.226	0.053	-0.001	0.272	0.179
	(0.187)	(0.171)	(0.175)	(0.162)	(0.186)	(0.177)	(0.175)	(0.161)	(0.178)	(0.174)
Treatment 2	0.387**	0.417**	0.248	0.301*	0.071	-0.017	0.109	0.096	0.126	0.017
	(0.181)	(0.167)	(0.166)	(0.155)	(0.181)	(0.175)	(0.174)	(0.165)	(0.178)	(0.173)
Treatment 3	0.138	0.077	-0.011	-0.056	0.095	-0.022	0.090	0.051	0.177	0.072
	(0.184)	(0.165)	(0.177)	(0.157)	(0.188)	(0.179)	(0.176)	(0.165)	(0.182)	(0.180)
Baseline likely use		0.309***		0.287***		0.254***		0.296***		0.173***
		(0.035)		(0.033)		(0.035)		(0.036)		(0.034)
Constant	5.641***	4.081***	5.449***	4.060***	4.880***	3.751***	5.428***	3.929***	4.667***	4.006***
	(0.225)	(0.286)	(0.207)	(0.258)	(0.219)	(0.265)	(0.214)	(0.292)	(0.226)	(0.261)
Observations	780	780	780	780	780	780	780	780	780	780
R-squared	0.136	0.246	0.132	0.238	0.117	0.188	0.107	0.209	0.091	0.128
p-value (T1≤C)	0.662	0.833	0.851	0.944	0.749	0.900	0.381	0.502	0.0635	0.151
p-value (T2≤T1)	0.003	0.000	0.003	0.000	0.138	0.115	0.361	0.255	0.808	0.842
p-value (T3≤T2)	0.934	0.985	0.944	0.990	0.446	0.510	0.547	0.614	0.385	0.371

Notes: Table reports results from two-sided test for treatment effects on the likelihood of future use of a series of a range of entities. Also reported are one-sided tests of pre-specified hypotheses for incremental positive treatment effects from each treatment arm compared against the preceding arm in the sequence. Regression includes vector of controls listed in Table 11. *** p<0.01, ** p<0.05, * p<0.1

against Treatment 1, beyond what can be achieved by a general warning message alone. Finally, we fail to reject the null hypothesis of no further incremental benefit associated with Treatment 3 when set against Treatment 2. It is possible that heightened likelihood of future use may reflect participants' increased confidence in filtering out fraudulent communications, enabling them to engage with confidence with relevant DFS providers.

4.1.2 Heterogeneous effects

We next examine the extent to which the impact of our learning interventions varied across subgroups of interest, including relevant experiential, behavioural, as well as demographic characteristics. Specifically, we test whether our treatment effects are more pronounced across levels of fraud, DFS, or ICT experience, self-control, risk appetite, or trust. We additionally test whether sector of employment (agriculture vs. non-agriculture), age, education, or gender significantly interacts with our interventions. We test for these effects across our four primary outcomes of interest: overall accuracy, true positives, true negatives, and confidence.

Do these experiential and demographic factors act as substitutes or complements to the anti-fraud campaign? That is, first, do we find differential performance in level terms across these subgroups in primary outcomes, and second, do particular subgroups respond more intensively to the educational interventions tested? We test these questions by exploring

heterogeneous treatment effects in overall accuracy, true positives, true negatives, and confidence. These results are reported in Tables 16-23).

For the most part, we find no strong evidence for significant level or interaction effects across these subgroups of interest. However, in some instances, we do find evidence of significant coefficient estimates in interactions across Tables 16-23. Mindful of the difficulty in direct interpretation of interaction effects in regression tables, to assist with interpretation, we take this subset of factors where we seem to observe significant differences, and depict the relevant relationships graphically in Figures 4-33.

First considering level differences in outcomes across relevant subgroups, we find evidence that DFS experience may act as a substitute for the learning interventions, with participants with high DFS experience performing better in overall accuracy than those with low experience (Figures 4 and 5). This is an intuitive result, likely reflecting an advantage conferred by experience in recognising what are plausible and more suspect communications. We find tentative evidence that lower trust levels may also partially substitute for the learning interventions, with generally higher performance in accuracy among those with low trust when compared against higher trust counterparts (see Figures 20-25). It is likely that lower trust participants show a higher degree of skepticism towards inbound communications purporting to originate with a genuine service provider, with a lower threshold for flagging fraud. This interpretation is supported by the fact that the higher accuracy for low trust participants is driven by a higher rate of true negatives, and not true positives. We additionally find that low trust participants report a higher confidence in level terms in their decisions than high trust counterparts. We observe some evidence in level terms across treatment arms that men report a higher degree of confidence in their judgements than women, and more tentative evidence that men have a higher level of overall accuracy than women (Figures 28-31). Those with high levels of self-control, and with higher risk appetite appear to demonstrate a higher level of confidence in their judgements over presented scenarios across treatment arms (Figures 26 and 27, and Figures 32 and 33).

Considering next how relevant characteristics may significantly interact with our treatment interventions, we find for the most part, no evidence of significant interaction effects across these experiential, behavioural, and demographic factors of interest. However, we do find isolated instances of apparent meaningful divergence in outcomes across subgroups, where significant interaction effects in treatment can be observed.

With the interpretative aide of graphical representation, it is clear that while we do observe isolated instances where coefficients associated with a given treatment arm do differ significantly within partitioned subgroups, largely, these outcomes do not meaningfully differ from their respective controls. As such, in these cases, we can say we do not find strong

Table 7: Learning by doing effects

	(1) Accuracy	(2) Confidence
Order	-0.003***	-0.009***
	(0.001)	(0.002)
Constant	0.643***	5.714***
	(0.008)	(0.045)
Observations	15,600	15,600
Number of participants	780	780
p-value ($\beta \leq 0$)	1	1

Note: Table reports results from two-sided test for 'learning by doing' effects on overall accuracy, and confidence, by regressing these outcomes on the order variable capturing the sequence of scenarios presented. Also reported are one-sided tests of hypotheses for positive learning effects. *** p<0.01, ** p<0.05, * p<0.1

or consistent evidence that certain subgroups respond more intensively to treatment than others.[28]

One tentative trend observed is an apparent convergence in confidence as treatments become more intense for certain subgroups that show an initial deficit. Lower risk appetite participants (Figure 32), females (Figure 30), and to a lesser extent low self-control (Figure 26) and low DFS experience participants (Figure 12) show signs of catch-up from an initial deficit in confidence when compared against their opposing indexed counterparts (i.e. high risk appetite, males, high self-control, and high DFS experience participants respectively). While the evidence in this regard is not strong, it is intuitive that those initially lacking in confidence in comparative terms may showe a higher marginal return in the confidence-engendering effect of targeted learning interventions.

[28]To take a case in point, we find in Table 16 (and corresponding Figures 4 and 5) a significant divergence in accuracy scores across those with and without prior fraud experience who were exposed to Treatment 1. However, in neither case does the outcome diverge significantly from the respective control condition, indicating the absence of a meaningful heterogeneity in treatment effects properly understood. Similar piecemeal patterns are observed widely in our data. As such, we do not claim to find strong evidence of heterogeneous treatment effects, notwithstanding the presence of some noteworthy level effects described previously.

4.1.3 Learning by doing

We assess additionally whether performance improves and confidence grows as participants proceed through scenarios presented as part of the experimental task. Specifically, whether those scenarios that appear later in the sequence are more accurately and confidently called. In Table 7, we find no evidence for such learning effects in accuracy or confidence, failing to reject the null hypothesis of no positive effect. Figures 2 and 3 graphically depict these outcomes over the sequence of scenarios presented. The absence of learning effects of this sort may reflect the degree of variety in scenarios presented, such that no systematic patterns or heuristics quickly establish themselves in the minds of participants.

4.1.4 Knowledge retention quiz

In Table 8, we test whether scores obtained by participants in a knowledge retention quiz designed to assess durability of the learning interventions varied by treatment group. In a poisson model, we estimate the impact from treatment on the count of correct answers reported in the quiz. We find modest evidence of heightened performance among treated participants. Participants who received any treatment perform only between 1.15 and 1.17 times better than control subjects in the count of correct answers.[29]

4.2 Results of UCC intervention

In Table 9, we report results from a final experimental manipulation, the UCC follow-up exercise. This tests the impact on the likelihood of user engagement of a unique pre-specified authentication code embedded in SMS communication. We find that the mere presence of a UCC code in communication does not significantly increase the likelihood of engagement. In addition, neither does the effectiveness of the code as a signal of authenticity enhanced when the recipient has personalised their own code, instead of having it automatically generated and assigned. In both cases we fail to reject the null hypothesis of no incremental impact on the likelihood of engagement.

[29] As part of the follow-up contact, experimental participants are asked about their hypothetical future preference regarding the format of authentication codes, considering the options of a numerical code, a word, or a sentence/phrase. 82% prefer a numerical code, 12% prefer a word, and 6% prefer a sentence/phrase. However, it is likely that the anchoring effect of recent experience with a numerical code over the course of their participation in the study influences these choices. As such, they should not be seen as clean or organic measures of preference.

Table 8: Impact of treatment on subsequent quiz score

	(1)
Treatment 1	0.159***
	(0.061)
Treatment 2	0.136**
	(0.061)
Treatment 3	0.137**
	(0.061)
Constant	0.514***
	(0.049)
Observations	519

Note: Table reports results from a poisson regression, estimating the number out of 3 quiz questions relating to key signs of fraud correctly answered in a follow-up call. Under a poisson specification, the expoentiated coefficient gives the multiplicative term with which to calculate the expected quiz score when the given treatment has been administered, relative to the control condition. As such, Treatment 1 participants are estimated to have $e^{0.159} = 1.17$ as a Rate Ratio, the mulciplicative increase in the expected quiz score compared to Control participants. For Treatment 2: $e^{0.136} = 1.15$. For Treatment 3: $e^{0.137} = 1.15$. *** p<0.01, ** p<0.05, * p<0.1

5 Conclusion

In this paper, we demonstrate the difficulty in meaningfully improving detection ability between genuine and fraudulent communications. We test three learning interventions, which vary in intensity from simple warning messages about the risk of fraud and the importance of vigilance, to a much deeper illustration of key signs of fraud with applied illustrative scenarios lasting 25 minutes. None of these treatments are successful in significantly improving performance in an experimental task where participants are asked to judge the authenticity or otherwise of 20 fictionalised communications scenarios. We fail to reject the null hypothesis of no incremental benefit from each treatment arm, in respect of overall accuracy, true positives, or true negatives. In this, we join the null effect column in the contested accounting over the remedial value of educational interventions in the financial and digital fraud domains. In view of the direct and specific nature of the learning interventions embodied in Treatments 2 and 3, the contextually-adjusted and engaging nature of the interventions, and the immediacy of the experimental task that followed, this failure is nonetheless surprising and disappointing. We do not find evidence of the returns to intensity in educational interventions observed elsewhere, and nor does the fact that interventions are delivered just in time, and at a relevant teachable moment yield the beneficial impact that the existing

Table 9: Impact of authentication on engagement probability

	(1)	(2)
UCC present	0.034	0.057
	(0.032)	(0.041)
UCC personalised		0.103**
		(0.043)
UCC present # UCC personalised		-0.049
		(0.063)
Constant	0.246***	0.198***
	(0.022)	(0.027)
Observations	780	780
R-squared	0.001	0.010
p-value ($\beta \leq 0$)	0.142	0.779

Note: Table reports from an OLS regression model predicting whether the participant responded to the outreach with the requested personal information. Column 1 includes a simple treatment condition recording when the UCC was embdedded in the SMS. Column 2 adds an interaction recording when that UCC was personalised, as distinct from having been automatically generated. *** p<0.01, ** p<0.05, * p<0.1

literature would lead us to hope for.

Our results cast some doubt over the promise of light-touch targeted educational interventions in moving the dial on stubborn and persistent vulnerabilities faced by users in navigating the digital financial landscape today. More immediately, our results highlight the severity of the challenge which MSEs in Nigeria are likely to face in avoiding the pitfalls presented by pervasive fraud.

One important caveat to register against these results is that the coefficients estimated for our treatment arms with respect to accuracy-related outcomes are all below our ex-post estimated minimum detectable effects, which range from 5% in overall accuracy, to 11% in true positives. As such, we cannot exclude the possibility that true effects from our learning intervention may have fallen in this range. However, the real-world economic meaning of treatment effects in this range, when set against the time cost of administering the learning intervention, and the fact that performance is tested immediately in a lab environment, with no competing demands on participants' attention, can be said to be limited.[30]

[30]However, the lack of ecological validity in our lab setting, where participants do not face real financial or other exposure from erroneous judgement of communications, may limit degree to which ours can be regarded as a definitive evaluation of the potential impact of these learning interventions.

Neither do we find evidence that a technical solution to the problem of authentication shows promise. In a separate experimental manipulation, the presence of a unique communications code embedded in follow-up communication with participants does not increase the likelihood of engagement, and neither does personalisation of the UCC strengthen its effectiveness as a signal of authenticity to increase engagement. While the failure to elicit positive treatment effects from the UCC is notable, it should be regarded as a less conclusive verdict than the failure of our learning interventions, where the experimental task represented a direct test of discriminant ability, and in a lab setting where achieving meaningful treatment effects should be comparatively easy. To illustrate the contrasting environments, in a follow-up phone survey, the most common reason cited among participants for non-response (both among those with and without the UCC embedded) was that participants were too busy, or simply forgot. Given the likely low real-world salience of the UCC follow-up task in the minds of participants, our UCC result should not be viewed as a final verdict on the technology.

While we do not find evidence of positive impacts from our learning interventions on the primary targeted objective in fraud detection performance, we do find that they produce effects on the degree of confidence which participants report in their judgements over presented scenarios. This result is troubling when set against the absence of any corresponding improvement in actual performance, and speaks to the risk associated with ineffectual learning interventions, where recipients may feel stronger in the given field, having undergone some targeted training, without actually being any stronger. That is to say, it speaks to the danger of false confidence effects. We do, however, find positive treatment effects on other associated outcomes: trust in DFS, the likelihood of future use of banks and mobile banking, and improved knowledge regarding the signs of fraud.

We find some tentative evidence of divergent patterns in performance across relevant subgroups in several areas. However, this evidence is subtle and secondary, and as such is not trumpeted loudly. We do not find evidence of significant heterogeneity in treatment effects across subgroups defined by relevant experiential, behavioural, and socio-demographic characteristics, but we do find some evidence of significant level effects in performance across treatment conditions. Intuitively, we observe that higher DFS experience, and lower trust participants are more accurate in their judgements over fictionalised scenarios than their less experienced and higher trust indexed counterparts. Another suggestive result relates to apparent catch-up growth in confidence levels reported among subgroups initially showing a deficit in confidence when compared against their indexed counterparts, as treatments increase in intensity. This subtle trend is evident for lower risk-appetite, female, lower self-control, and lower DFS-experienced participants, pointing to a potentially higher marginal

return in engendered confidence from learning interventions.

Notwithstanding the failure to yield encouraging results in the improvement of fraud detection, there remains great urgency in devising effective interventions to combat digital fraud due to the large and growing costs associated with digital deception. Future research should not be discouraged from this challenge, but integrate the lessons from past efforts, taking particular cautionary note of associated risks.[31]

[31]Instructive lessons will be taken forward from the results of this lab trial to inform the design of a future field trial aimed at combating susceptibility to digital fraud in the real-world, using a sample of new and existing users of the Amana Market platform in Nigeria.

References

Akerlof, Geokge A, "The market for lemons: quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Anderson, KB, "Off the hook: Reducing participation in telemarketing fraud," *Washington, DC: AARP Foundation*, 2003.

Archana, BS, Ashika Chandrashekar, Anusha Govind Bangi, BM Sanjana, and Syed Akram, "Survey on usable and secure two-factor authentication," in "2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)" IEEE 2017, pp. 842–846.

Ba, Bocar, "Going the extra mile: The cost of complaint filing, accountability, and law enforcement outcomes in Chicago," Technical Report, Working paper 2018.

Bandura, Albert, "Self-efficacy: toward a unifying theory of behavioral change.," *Psychological review*, 1977, *84* (2), 191.

_ , William H Freeman, and Richard Lightsey, "Self-efficacy: The exercise of control," 1999.

Banerjee, Abhijit, Emily Breza, Esther Duflo, and Cynthia Kinnan, "Can microfinance unlock a poverty trap for some entrepreneurs?," Technical Report, National Bureau of Economic Research 2019.

_ , _ , _ , and _ , "Can Microfinance Unlock a Poverty Trap for Some Entrepreneuers?," 2021.

Barson, P, S Field, N Davey, G McAskie, and R Frank, "The detection of fraud in mobile phone networks," *Neural Network World*, 1996, *6* (4), 477–484.

Bartlett, Frederic Charles and Frederic C Bartlett, *Remembering: A study in experimental and social psychology*, Cambridge University Press, 1995.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey, "Managerial Miscalibration," *The Quarterly Journal of Economics*, November 2013, *128* (4), 1547–1584.

Bird, Matthew and Rafe Mazer, "Uganda Consumer Protection in Digital Finance Survey," Technical Report 2021.

Biros, David P, Joey F George, and Robert W Zmud, "Inducing sensitivity to deception in order to improve decision making performance: A field study," *MIS quarterly*, 2002, pp. 119–144.

Blackmon, William, Rafe Mazer, and Shana Warren, "Kenya Consumer Protection in Digital Finance Survey," Technical Report 2021.

_ , _ , and _ , "Nigeria Consumer Protection in Digital Finance Survey," Technical Report 2021.

Boonkrong, Sirapat, *Authentication and Access Control: Practical Cryptography Methods and Tools*, Springer, 2021.

Breza, Emily and Cynthia Kinnan, "Measuring the Equilibrium Impacts of Credit: Evidence from the Indian Microfinance Crisis," *Quarterly Journal of Economics*, 2021, *136* (3), 1447–1497.

Burgoon, Judee K, "Deception detection accuracy," *The International Encyclopedia of Interpersonal Communication*, 2015, pp. 1–6.

_ , David B Buller, Amy S Ebesu, and Patricia Rockwell, "Interpersonal deception: V. Accuracy in deception detection," *Communications Monographs*, 1994, *61* (4), 303–325.

Burke, Jeremy, Christine N Kieffer, Gary R Mottola, and Francisco Pérez-Arce, "Can Educational Interventions Reduce Susceptibility to Financial Fraud?," *Available at SSRN 3747165*, 2020.

Byrne, Shane, Daniel Putman, Michael King, and Chaning Jang, "Strategies for Reducing Non-Institutional Fraud and Building Trust in a Digital Market Platform: A Behavioral Lab Experiment in Nigeria," *AEA RCT Registry*, 2022.

Caribou Digital, "Returns to Investment in Higher Education," 2019.

Conroy, Julie, "Digital Authentication: New Opportunities to Enhance the Customer Journey," *Aite Group, Inc. https://aite-novarica.com/report/digital-authentication-new-opportunities-enhance-customer-journey*, 2017.

Darby, Michael R and Edi Karni, "Free competition and the optimal amount of fraud," *The Journal of law and economics*, 1973, *16* (1), 67–88.

DeLiema, Marguerite, Gary R Mottola, and Martha Deevy, "Findings from a pilot study to measure financial fraud in the United States," *Available at SSRN 2914560*, 2017.

DeLiema, Marti, Emma Fletcher, Christine Kleffer, Gary Mottola, Rubens Pessanha, and Melissa Trumpower, "Exposed to Scams: What Separates Victims from Non-Victims?," 2019.

Dong, Anmei, Morris Siu-Yung Jong, and Ronnel B King, "How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking," *Frontiers in psychology*, 2020, *11*, 591203.

Elbers, Chris, Jan Willem Gunning, and Bill Kinsey, "Growth and Risk: Methodology and Micro Evidence," *World Bank Economic Review*, 2007, *21* (1), 1–20.

Engels, Christian, Kamlesh Kumar, and Dennis Philip, "Financial literacy and fraud detection," *The European Journal of Finance*, 2020, *26* (4-5), 420–442.

Erel, Isil and Jack Liebersohn, "Can FinTech Reduce Disparities in Access to Finance? Evidence from the Paycheck Protection Program," *Journal of Financial Economics*, October 2022, *146* (1), 90–118.

Fernandes, Daniel, John G Lynch Jr, and Richard G Netemeyer, "Financial literacy, financial education, and downstream financial behaviors," *Management Science*, 2014, *60* (8), 1861–1883.

Fisch, Jill E and Jason S Seligman, "Trust, financial literacy, and financial market participation," *Journal of Pension Economics & Finance*, 2022, *21* (4), 634–664.

Florêncio, Dinei and Cormac Herley, "Sex, lies and cyber-crime surveys," in "Economics of information security and privacy III," Springer, 2013, pp. 35–53.

Francois, Patrick and Jan Zabojnik, "Trust, social capital, and economic development," *Journal of the European Economic Association*, 2005, *3* (1), 51–94.

Fu, Jonathan and Mrinal Mishra, "Combatting fraudulent and predatory fintech apps with machine learning," 2022.

Garz, Seth, Xavier Giné, Dean Karlan, Rafe Mazer, Caitlin Sanford, and Jonathan Zinman, "Consumer protection for financial inclusion in low-and middle-income countries: Bridging regulator and academic perspectives," *Annual Review of Financial Economics*, 2021, *13*, 219–246.

George, Joey F, Kent Marett, and Patti Tilley, "Deception detection under varying electronic media and warning conditions," in "37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the" IEEE 2004, pp. 9–pp.

Grazioli, Stefano and Alex Wang, "Looking without seeing: understanding unsophisticated consumers' success and failure to detect Internet deception," 2001.

Griffin, John M., Samuel Kruger, and Prateek Mahajan, "Did FinTech Lenders Facilitate PPP Fraud?," *The Journal of Finance*, June 2023, *78* (3), 1777–1827.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Trusting the stock market," *The Journal of Finance*, 2008, *63* (6), 2557–2600.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, "Trust Busting: The Effect of Fraud on Investor Behavior," *The Review of Financial Studies*, April 2018, *31* (4), 1341–1376.

Harrison, Glenn W and John A List, "Field Experiments," *Journal of Economic Literature*, 2004.

Hartwig, Maria and Charles F Bond Jr, "Why do lie-catchers fail? A lens model meta-analysis of human lie judgments.," *Psychological bulletin*, 2011, *137* (4), 643.

Herley, Cormac, "So long, and no thanks for the externalities: the rational rejection of security advice by users," in "Proceedings of the 2009 workshop on new security paradigms" 2009, pp. 133–144.

_ , "Why do nigerian scammers say they are from nigeria?," in "WEIS" Berlin 2012.

Hilal, Waleed, S Andrew Gadsden, and John Yawney, "Financial Fraud:: A Review of Anomaly Detection Techniques and Recent Advances," 2022.

Holtfreter, Kristy, Michael D Reisig, Nicole Leeper Piquero, and Alex R Piquero, "Low self-control and fraud: Offending, victimization, and their overlap," *Criminal Justice and Behavior*, 2010, *37* (2), 188–203.

Innovations for Poverty Action, "Consumer Protection in Digital Finance Surveys," `https://www.poverty-action.org/consumer-protection-digital-finance-surveys` 2021.

_ , "Social media usage by digital finance consumers: Analysis of consumer complaints in Kenya, Nigeria and Uganda. July 2019 - July 2020," `https://www.poverty-action.org/sites/default/files/publications/Social-Media-Usage-by-Digital-Finance-Consumers-April-2021.pdf` 2021.

Jack, William and Tavneet Suri, "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution," *American Economic Review*, 2014, *104* (1), 183–223.

Kabanda, Salah, Maureen Tanner, and Cameron Kent, "Exploring SME cybersecurity practices in developing countries," *Journal of Organizational Computing and Electronic Commerce*, 2018, *28* (3), 269–282.

Kaiser, Tim and Lukas Menkhoff, "Does financial education impact financial literacy and financial behavior, and if so, when?," *The World Bank Economic Review*, 2017, *31* (3), 611–630.

_ **, Annamaria Lusardi, Lukas Menkhoff, and Carly Urban**, "Financial education affects financial knowledge and downstream behaviors," *Journal of Financial Economics*, 2021.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry, "Agricultural Decisions After Relaxing Credit and Risk Constraints," *The Quarterly Journal of Economics*, 2014, *129* (2), 597–652.

Kester, Liesbeth, Gemma Corbalan, and Femke Kirschner, "Cognitive load theory and multimedia learning, task characteristics, and learning engagement: The current state of the art," *Computers in Human Behavior*, 2011, *27* (1), 1–4.

Kirkos, Efstathios, Charalambos Spathis, and Yannis Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," *Expert systems with applications*, 2007, *32* (4), 995–1003.

Kraft, Matthew A, "Interpreting effect sizes of education interventions," *Educational Researcher*, 2020, *49* (4), 241–253.

Kubilay, Elif, Eva Raiber, Lisa Spantig, Lucy Kaaria, and Cahlíková Jana, "Can You Spot A Scam? Measuring and Improving Scam Identification Ability," 2023.

Lusardi, Annamaria, ""Just in time education"? Just in time is too late," *Wall Street Journal*, 2015.

Lyons, Benjamin A., Jacob M. Montgomery, Andrew M. Guess, Brendan Nyhan, and Jason Reifler, "Overconfidence in News Judgments Is Associated with False News Susceptibility," *Proceedings of the National Academy of Sciences*, June 2021, *118* (23), e2019527118.

Mazer, Rafe and Matthew Bird, "Consumer Protection Survey of Digital Finance Users: Uganda," 2021.

_ **and Shana Warren**, "Consumer Protection Survey of Digital Finance Users: Kenya," 2021.

McKenzie, David, "Tools of the Trade: a joint test of orthogonality when testing for balance," *World Bank Development Impact Blog*, 2015.

_ **and Owen Ozier**, "Why ex-post power using estimated effect sizes is bad, but an ex-post MDE is not," *World Bank Development Impact Blog*, 2019.

Meager, Rachael, "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, 2019, *11* (1), 57–91.

Miles, Stan and Derek Pyne, "The economics of scams," *Review of Law & Economics*, 2017, *13* (1).

Miller, Margaret, Julia Reichelstein, Christian Salas, and Bilal Zia, "Can you help someone become financially capable? A meta-analysis of the literature," *The World Bank Research Observer*, 2015, *30* (2), 220–246.

Mutz, Diana C, Robin Pemantle, and Philip Pham, "The perils of balance testing in experimental design: Messy analyses of clean data," *The American Statistician*, 2019, *73* (1), 32–42.

Nash, Rebecca, Martin Bouchard, and Aili Malm, "Investing in people: The role of social networks in the diffusion of a large-scale fraud," *Social networks*, 2013, *35* (4), 686–698.

OECD, "G20/OECD INFE Report: Ensuring financial education and consumer protection for all in the digital age," 2017.

Raval, Devesh, "Whose voice do we hear in the marketplace? Evidence from consumer complaining behavior," *Marketing Science*, 2020, *39* (1), 168–187.

Riley, Emma, "Mobile Money and Risk Sharing against Village Shocks," *Journal of Development Economics*, 2018, *135* (June), 43–58.

_ **and Abu Shonchoy**, "A National Information Campaign Encouraging Financial Technology Use in Ghana," 2022.

Sahin, Yusuf and Ekrem Duman, "Detecting credit card fraud by decision trees and support vector machines," in "World Congress on Engineering 2012. July 4-6, 2012. London, UK.," Vol. 2188 International Association of Engineers 2010, pp. 442–447.

Scheibe, Susanne, Nanna Notthoff, Josephine Menkin, Lee Ross, Doug Shadel, Martha Deevy, and Laura L Carstensen, "Forewarning reduces fraud susceptibility in vulnerable consumers," *Basic and applied social psychology*, 2014, *36* (3), 272–279.

Scott, Elizabeth D, "Just (?) a true-false test: Applying signal detection theory to judgments of organizational dishonesty," *Business & society*, 2006, *45* (2), 130–148.

Serra-Garcia, Marta and Uri Gneezy, "Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies," *American Economic Review*, October 2021, *111* (10), 3160–3183.

Stankov, Lazar and John D Crawford, "Confidence judgments in studies of individual differences," *Personality and Individual Differences*, 1996, *21* (6), 971–986.

Statman, Meir, Steven Thorley, and Keith Vorkink, "Investor Overconfidence and Trading Volume," *Review of Financial Studies*, 2006, *19* (4), 1531–1565.

Stiglitz, Joseph E and Andrew Weiss, "Credit rationing in markets with imperfect information," *The American Economic Review*, 1981, *71* (3), 393–410.

Stone-Gross, Brett, Ryan Abman, Richard A Kemmerer, Christopher Kruegel, Douglas G Steigerwald, and Giovanni Vigna, "The underground economy of fake antivirus software," in "Economics of information security and privacy III," Springer, 2013, pp. 55–78.

Suri, Tavneet, Prashant Bharadwaj, and William Jack, "Fintech and Household Resilience to Shocks: Evidence from Digital Loans in Kenya," *Journal of Development Economics*, 2021, *153* (April 2020), 102697.

Sweller, John, Jeroen JG Van Merrienboer, and Fred GWC Paas, "Cognitive architecture and instructional design," *Educational Psychology Review*, 1998, *10* (3), 251–296.

Tade, Oludayo, "COVID-'419': Social Context of Cybercrime in the Age of COVID-19 in Nigeria," *African Security*, 2021, *14* (4), 460–483.

Titus, Richard M, Fred Heinzelmann, and John M Boyle, "Victimization of persons by fraud," *Crime & Delinquency*, 1995, *41* (1), 54–72.

Vancouver, Jeffrey B and Laura N Kendall, "When self-efficacy negatively relates to motivation and performance in a learning context.," *Journal of Applied Psychology*, 2006, *91* (5), 1146.

_ , Charles M Thompson, and Amy A Williams, "The changing signs in the relationships among self-efficacy, personal goals, and performance.," *Journal of Applied Psychology*, 2001, *86* (4), 605.

Velásquez, Ignacio, Angélica Caro, and Alfonso Rodríguez, "Authentication schemes and methods: A systematic literature review," *Information and Software Technology*, 2018, *94*, 30–37.

Vohs, Kathleen D, Roy F Baumeister, and Jason Chin, "Feeling duped: Emotional, motivational, and cognitive aspects of being exploited by others," *Review of General Psychology*, 2007, *11* (2), 127–141.

Walters, Daniel J. and Philip M. Fernbach, "Investor Memory of Past Performance Is Positively Biased and Predicts Overconfidence," *Proceedings of the National Academy of Sciences*, September 2021, *118* (36), e2026680118.

Wang, Yibo and Wei Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud," *Decision Support Systems*, 2018, *105*, 87–95.

Warren, Shana and Rafe Mazer, "Consumer Protection Survey of Digital Finance Users: Nigeria," 2021.

Woodman, TIM and LEW Hardy, "The relative impact of cognitive anxiety and self-confidence upon sport performance: A meta-analysis," *Journal of Sports Sciences*, 2003, *21* (6), 443–457.

Woodman, Tim, Sally Akehurst, Lew Hardy, and Stuart Beattie, "Self-confidence and performance: A little self-doubt helps," *Psychology of Sport and Exercise*, 2010, *11* (6), 467–470.

World Bank Group, "Nigeria Digital Economy Diagnostic Report," 2019.

Xiao, Bo and Izak Benbasat, "Designing warning messages for detecting biased online product recommendations: An empirical investigation," *Information Systems Research*, 2015, *26* (4), 793–811.

Yeh, Ting-Kuang, Kuan-Yun Tseng, Chung-Wen Cho, James P Barufaldi, Mei-Shin Lin, and Chun-Yen Chang, "Exploring the impact of prior knowledge and appropriate feedback on students' perceived cognitive load and learning outcomes: Animation-based earthquakes instruction," *International Journal of Science Education*, 2012, *34* (10), 1555–1570.

Youll, Jim, "Fraud vulnerabilities in Sitekey security at Bank of America," *Available: www. cr-labs. com/publications/SiteKey-20060718. pdf*, 2006.

Appendix

A Empirical methodology - extended

A.1 Standard error adjustments

Treatment assignment is at the individual level, therefore for outcomes with multiple observations per participant, we will apply cluster robust standard errors at the individual level. For any outcomes with only one observation per treatment unit, we will apply heteroskedasticity robust standard errors.

A.2 Multiple hypothesis testing

As described in the sections above, we opt to reduce the number of tests in each outcome group as opposed to adjusting for multiple hypothesis testing. Specifically, we test a single outcome for each primary outcome group. Where multiple outcomes are of interest, we will construct a standardized index of the outcomes to serve as the primary outcome for that group as in Anderson (2008).

A.3 Statistical power

To assess the sample size requirement for the experiment, we estimate the minimum detectable effects (MDE) under a range of alternative design scenarios (varying by sample size, power, and number of treatment arms) in Table 10.

The table provides an estimate of the smallest treatment effect that could be detected with statistical confidence were it to be achieved by one of our educational interventions. For a given scenario, treatment effects smaller than that reported would not be detectable with statistical confidence. The results in the table are calculated relative to a baseline unconditional probability of unaided detection ability of 50% (i.e. chance).[32]

For example, starting with the top left scenario, if only one treatment were administered and the number of participants was 250, it would not be possible to detect treatment effects smaller than a 20.58% improvement over the baseline detection accuracy (with 90% power). The power level refers to an acceptable level of probability that the experiment will detect an effect when the effect is present. In this example, if we were to repeat the experiment over and over, we would detect an impact at least as big as this 90% of the time.

[32]Unaided human deception detection capacity has been estimated to be little better than chance (Hartwig and Bond Jr, 2011).

Ex-ante, in our chosen design, our estimated MDE with $N = 780$ and 3 treatment arms was 24.90% (with 80% power), or 29.3% (with 90% power). Following McKenzie and Ozier (2019), we calculate ex-post MDEs using actual realised data from the control group. Ex-post, using the observed accuracy and standard deviation in the control group ($\mu = 0.61$, $\sigma = 0.13$), we estimate our MDE for treatment effects in accuracy to have been much lower, at 5.24% (with 80% power), or 6.16% (with 90% power). In true positives, with observed outcomes in the control group ($\mu = 0.55$, $\sigma = 0.25$), our MDE is 11.28% (with 80% power), or 13.27% (with 90% power). In true negatives, with observed outcomes in the control group ($\mu = 0.67$, $\sigma = 0.24$), our MDE is 9.15% (with 80% power), or 10.76% (with 90% power). Tr.: number of treatment arms.

Table 10: Estimated statistical power under alternative design scenarios

Outcome: detection accuracy									
	N	Power 90%				Power 80%			
Tr.		1	2	3	4	1	2	3	4
	250	31.54%	38.76%	44.90%	50.08%	37.12%	45.62%	52.86%	58.94%
	500	22.26%	27.34%	31.54%	35.28%	26.22%	32.18%	37.12%	41.52%
	750	18.18%	22.26%	25.76%	28.78%	21.40%	26.22%	30.32%	33.86%
	780	17.60%	21.58%	24.90%	27.86%	20.72%	25.40%	29.30%	32.78%
	1000	15.74%	19.28%	22.26%	24.90%	18.52%	22.68%	26.22%	29.32%

Notes: Table reports ex-ante power calculations giving minimum detectable effect sizes under a range of sample size and treatment arm scenarios. We assumed baseline accuracy of chance ($\mu = 0.50$, $\sigma = 0.50$). Each scenario estimates the minimum detectable effect, expressed as a percentage increase over baseline scores.

B Departures from pre-analysis plan

The preregistration for this experiment was filed with the American Economic Association's registry for randomised controlled trials in July 2022, before data collection began (Byrne et al., 2022). We depart from the research plan pre-specified in the pre-analysis plan (PAP) in a number of areas.

B.1 Pre-specified but not included

- PAP hypotheses 6.1 and 6.2 undertook to investigate how urgency (i.e. time pressure) in scenarios affected confidence and accuracy. Due to challenges in data reporting relating to the time limit imposed in different scenarios, it was not possible to perform these tests.

- PAP hypotheses 7.1 and 7.2 undertook to assess knowledge retention from our learning interventions, by testing for decay in performance between two time points: a quiz administered at the close of the endline survey, and a follow-up quiz administered at +3 weeks. We additionally undertook how this rate of decay varied in accordance with the intensity of the original treatment administered. Due to oversight in implementation, the first quiz was not administered as part of the endline survey, making it impossible to perform the pre-specified tests of decay. As an alternative, we test instead how performance in the follow-up quiz alone varies by treatment.

- Our PAP envisaged that our participant sample would be comprised partly of a supplemental pool of students from the partnering university in Nigeria, creating an occupational subcategory of 'student'. However, the opportunity to supplement our sample instead with additional users from the Amana Market platform arose. This was deemed preferable, corresponding more closely to the target population of interest.

B.2 Included but not pre-specified

- We pre-specified that we would examine learning by doing effects within the experiment, testing whether if those scenarios that appeared later in the order were more often correctly identified by participants. We did not additionally specify that we would examine whether confidence grows through the sequence of presented scenarios, but this outcome is also evaluated in the paper. Confidence is, however, pre-specified as one of our primary outcomes related to the ability to distinguish between genuine and

fraudulent communications. We viewed this as sufficiently interesting and important to merit inclusion.

C Additional tables and figures

Table 11: Variables of interest for heterogeneous treatment effects

Variable of interest	Detail
ICT experience	Standardized index of experiences with information communication technologies. After indexing, individuals are split into high and low experience types.
DFS experience	Standardized index of experiences with digital financial services. After indexing, individuals are split into high and low experience types.
Fraud experience	Participants are split into four types: those who have not encountered fraud, those who encountered fraud but did not respond, those who responded but did not suffer losses, and those who responded and suffered losses.
Gender	An indicator variable equal to one if the business owner is a woman, zero otherwise.
Occupation	A set of indicator variables (and a left-out group) for the following occupations: Agriculture, Non-Agriculture
Self-Control	A standardized index of self-control, impulsiveness, attentiveness. After indexing, individuals are split into those who have above or below average self control.
Risk Preference	A standardized index of risk preferences built from two question: a simple elicitation of risk preferences and a self-reported assessment of risk preferences. After indexing this may be split into high and low risk types.
Generalized Trust and Skepticism	A standardized index of variables associated with generalized trust and scepticism, including questioning mind. After indexing, participants are split into a high and low trust types.

Note: See Table 28 for complete definition of all variables used.

Table 12: Research hypotheses for core research questions

Research Question	No.	Hypothesis
Do anti-fraud interventions increase the ability to distinguish between fraud and legitimate communications?	1.0	Providing MSEs with the anti-fraud campaign improves their ability to distinguish between genuine and same fraudulent communications (T1, T2, and T3 vs. C).
	1.1	Providing MSEs with a general warning message about fraud alone (with no further educational intervention) improves their ability to distinguish between genuine and fraudulent communications (T1 vs. C).
	1.2	Providing MSEs with warning signs for potential fraud in a simple format improves their ability to distinguish between genuine and fraudulent communications, still further than can be achieved by a general warning message alone (T2 vs. T1).
	1.3	Illustrating applied examples of fraudulent communications in a simple format improves MSEs' ability to distinguish between genuine and fraudulent communications, still further than can be achieved with simple warning signs alone (T3 vs. T2).
Do anti-fraud interventions increase confidence in the ability to distinguish between fraud and legitimate communications?	2.0	Providing MSEs with the anti-fraud campaign improves their confidence in their ability to distinguish between fraudulent and legitimate communications.
Do anti-fraud interventions increase trust in digital financial services?	3.0	Providing MSEs with the anti-fraud campaign improves their trust in DFS.
Does a simple anti-fraud intervention increase usage of digital financial services?	4.0	Providing MSEs with the anti-fraud campaign improves their likelihood of using DFS in the future.
Is the UCC suitably deployed?	5.1	How does the presence (absence) of a pre-specified authentication code affect the degree of confidence recipients place in customer outreach?
	5.2	Is the effectiveness of a pre-specified authentication code as a signal of authenticity enhanced when the recipient has specified their own code, as against when it is automatically generated and assigned?
Is knowledge from educational interventions effectively retained over a short time horizon?	6.0	How does performance in a knowledge retention quiz relating to key signs of fraud administered at +3 weeks vary in accordance with the intensity of the original educational intervention administered?
Do participants learn by doing?	7.0	Does accuracy improve and confidence grow in respect of scenarios presented later in the sequence when compared against those at the start?

Table 13: Covariate balance by treatment

	(1) Treatment 1	(2) Treatment 2	(3) Treatment 3
Age	-0.007	-0.008*	0.005
	(0.005)	(0.005)	(0.005)
Female	-0.239***	-0.221***	-0.174***
	(0.054)	(0.055)	(0.055)
Third level education	0.038	0.010	0.016
	(0.054)	(0.056)	(0.055)
Married	0.065	0.018	0.006
	(0.061)	(0.062)	(0.063)
Agricultural employment	0.026	0.062	0.011
	(0.055)	(0.055)	(0.056)
Contacted by scammer	-0.001	0.063	0.085
	(0.056)	(0.057)	(0.059)
Access to smartphone	-0.110	0.155	-0.022
	(0.093)	(0.120)	(0.104)
Business owner	-0.011	0.019	-0.101
	(0.081)	(0.083)	(0.080)
Has formal financial account	-0.052	0.068	-0.124*
	(0.070)	(0.076)	(0.074)
Used online platforms	-0.001	0.031	0.061
	(0.055)	(0.054)	(0.055)
Trusting	0.095*	0.060	0.079
	(0.051)	(0.052)	(0.052)
Risk averse	-0.089	0.004	-0.052
	(0.055)	(0.055)	(0.056)
Constant	0.873***	0.492**	0.544***
	(0.177)	(0.195)	(0.174)
Observations	390	390	390
R-squared	0.083	0.074	0.066

Notes: Table reports linear prediction of treatment status (for each treatment arm) compared against the control group, across a range of important descriptive characteristics. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 14: Overall effect: partialling out lasso approach

	(1) Overall	(2) True positive	(3) True negative
Treatment 1	-0.010	-0.021	0.001
	(0.013)	(0.026)	(0.025)
Treatment 2	-0.001	-0.025	0.023
	(0.012)	(0.026)	(0.023)
Treatment 3	-0.009	-0.049*	0.032
	(0.012)	(0.026)	(0.023)
Observations	780	780	780

Notes: Table reports an the same effects reported in Table 3, but using an alternative approach to the selection of relevant control variables for the purpose of robustness: partialling out lasso linear regression. *** p<0.01, ** p<0.05, * p<0.1

Table 15: Confidence effect: partialling out lasso approach

	(1) Overall	(2) True positive	(3) True negative
Treatment 1	0.005	-0.010	0.017
	(0.114)	(0.115)	(0.119)
Treatment 2	0.207*	0.200*	0.214*
	(0.111)	(0.115)	(0.113)
Treatment 3	0.255**	0.206*	0.283**
	(0.114)	(0.117)	(0.117)
Observations	780	780	780

Notes: Table reports an the same effects reported in Table 4, but using an alternative approach to the selection of relevant control variables for the purpose of robustness: partialling out lasso linear regression. *** p<0.01, ** p<0.05, * p<0.1

Table 16: Heterogeneous effects in accuracy: experience and behaviour

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.005	-0.022	-0.002	0.008	0.000	-0.001
	(0.017)	(0.017)	(0.016)	(0.018)	(0.019)	(0.018)
Treatment 2	-0.003	-0.012	0.002	0.005	0.005	-0.006
	(0.016)	(0.019)	(0.016)	(0.018)	(0.019)	(0.017)
Treatment 3	-0.010	-0.022	-0.004	-0.005	-0.029	-0.005
	(0.016)	(0.017)	(0.016)	(0.018)	(0.018)	(0.017)
No direct experience	-0.014					
	(0.018)					
Treatment 1# No direct experience	-0.042*					
	(0.025)					
Treatment 2# No direct experience	0.012					
	(0.025)					
Treatment 3# No direct experience	0.008					
	(0.025)					
Low self-control		-0.029				
		(0.018)				
Treatment 1# Low self-control		0.025				
		(0.025)				
Treatment 2# Low self-control		0.021				
		(0.024)				
Treatment 3# Low self-control		0.027				
		(0.025)				
Low risk aversion			0.012			
			(0.017)			
Treatment 1#Low risk aversion			-0.030			
			(0.026)			
Treatment 2# Low risk aversion			-0.009			
			(0.024)			
Treatment 3# Low risk aversion			-0.013			
			(0.026)			
Low trust				0.037**		
				(0.018)		
Treatment 1# Low trust				-0.033		
				(0.025)		
Treatment 2# Low trust				-0.013		
				(0.024)		
Treatment 3# Low trust				-0.008		
				(0.025)		
Low DFS experience					-0.038**	
					(0.018)	
Treatment 1# Low DFS experience					-0.020	
					(0.025)	
Treatment 2# Low DFS experience					-0.012	
					(0.024)	
Treatment 3# Low DFS experience					0.045*	
					(0.025)	
Low ICT experience						-0.007
						(0.017)
Treatment 1# Low ICT experience						-0.016
						(0.025)
Treatment 2# Low ICT experience						0.013
						(0.025)
Treatment 3# Low ICT experience						-0.008
						(0.025)
Constant	0.652***	0.664***	0.650***	0.648***	0.655***	0.652***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)
Observations	780	780	780	780	780	780
R-squared	0.091	0.087	0.086	0.087	0.095	0.087

Notes: Table explores heterogeneous treatment effects in accuracy across experiential and behavioural characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 17: Heterogeneous effects in accuracy: demographic

	(1)	(2)	(3)	(4)
Treatment 1	-0.006	-0.011	0.001	-0.018
	(0.015)	(0.016)	(0.020)	(0.020)
Treatment 2	-0.002	-0.008	0.009	-0.011
	(0.014)	(0.015)	(0.019)	(0.019)
Treatment 3	-0.002	0.001	-0.009	-0.010
	(0.015)	(0.017)	(0.019)	(0.020)
Agriculture	0.001			
	(0.020)			
Treatment 1#Agriculture	-0.012			
	(0.027)			
Treatment 2#Agriculture	0.001			
	(0.027)			
Treatment 3# Agriculture	-0.019			
	(0.027)			
Above median age		0.006		
		(0.018)		
Treatment 1# Above median age		0.002		
		(0.025)		
Treatment 2# Above median age		0.017		
		(0.025)		
Treatment 3# Above median age		-0.016		
		(0.025)		
Lower education			-0.020	
			(0.018)	
Treatment 1# Lower education			-0.017	
			(0.025)	
Treatment 2# Lower education			-0.019	
			(0.025)	
Treatment 3# Lower education			0.002	
			(0.025)	
Female				-0.040**
				(0.019)
Treatment 1#Female				0.014
				(0.026)
Treatment 2# Female				0.019
				(0.026)
Treatment 3# Female				-0.000
				(0.025)
Constant	0.651***	0.654***	0.666***	0.659***
	(0.017)	(0.017)	(0.020)	(0.020)
Observations	780	780	780	780
R-squared	0.086	0.087	0.097	0.086

Notes: Table explores heterogeneous treatment effects in accuracy across demographic characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 18: Heterogeneous effects in true positives: experience and behaviour

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	-0.002	-0.032	-0.014	-0.015	0.007	-0.016
	(0.032)	(0.032)	(0.031)	(0.040)	(0.037)	(0.039)
Treatment 2	-0.058*	-0.027	-0.044	-0.018	-0.036	-0.048
	(0.031)	(0.037)	(0.033)	(0.038)	(0.038)	(0.036)
Treatment 3	-0.087***	-0.046	-0.053*	-0.047	-0.081**	-0.044
	(0.032)	(0.033)	(0.032)	(0.039)	(0.037)	(0.038)
No direct experience	-0.068*					
	(0.036)					
Treatment 1# No direct experience	-0.058					
	(0.055)					
Treatment 2# No direct experience	0.110**					
	(0.055)					
Treatment 3# No direct experience	0.129**					
	(0.054)					
Low self-control		-0.051				
		(0.035)				
Treatment 1# Low self-control		0.026				
		(0.053)				
Treatment 2# Low self-control		0.003				
		(0.050)				
Treatment 3# Low self-control		-0.006				
		(0.051)				
Low risk aversion			-0.016			
			(0.037)			
Treatment 1#Low risk aversion			-0.033			
			(0.057)			
Treatment 2# Low risk aversion			0.055			
			(0.051)			
Treatment 3# Low risk aversion			0.012			
			(0.056)			
Low trust				0.017		
				(0.035)		
Treatment 1# Low trust				-0.012		
				(0.052)		
Treatment 2# Low trust				-0.015		
				(0.050)		
Treatment 3# Low trust				-0.005		
				(0.052)		
Low DFS experience					-0.030	
					(0.035)	
Treatment 1# Low DFS experience					-0.053	
					(0.052)	
Treatment 2# Low DFS experience					0.022	
					(0.050)	
Treatment 3# Low DFS experience					0.071	
					(0.052)	
Low ICT experience						-0.003
						(0.036)
Treatment 1# Low ICT experience						-0.008
						(0.051)
Treatment 2# Low ICT experience						0.054
						(0.051)
Treatment 3# Low ICT experience						-0.013
						(0.052)
Constant	0.605***	0.594***	0.595***	0.587***	0.595***	0.594***
	(0.035)	(0.034)	(0.035)	(0.037)	(0.037)	(0.038)
Observations	780	780	780	780	780	780
R-squared	0.039	0.020	0.023	0.020	0.027	0.022

Notes: Table explores heterogeneous treatment effects in true positives across experiential and behavioural characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 19: Heterogeneous effects in true positives: demographic

	(1)	(2)	(3)	(4)
Treatment 1	0.007	-0.036	-0.014	-0.050
	(0.032)	(0.034)	(0.037)	(0.038)
Treatment 2	-0.001	-0.001	-0.014	-0.061
	(0.033)	(0.033)	(0.036)	(0.038)
Treatment 3	-0.004	-0.013	-0.077**	-0.081**
	(0.032)	(0.039)	(0.036)	(0.039)
Agriculture	0.073**			
	(0.037)			
Treatment 1#Agriculture	-0.091*			
	(0.054)			
Treatment 2#Agriculture	-0.080			
	(0.052)			
Treatment 3# Agriculture	-0.135**			
	(0.053)			
Above median age		0.029		
		(0.035)		
Treatment 1# Above median age		0.032		
		(0.052)		
Treatment 2# Above median age		-0.055		
		(0.051)		
Treatment 3# Above median age		-0.069		
		(0.053)		
Lower education			0.011	
			(0.035)	
Treatment 1# Lower education			-0.015	
			(0.051)	
Treatment 2# Lower education			-0.022	
			(0.050)	
Treatment 3# Lower education			0.053	
			(0.051)	
Female				-0.047
				(0.037)
Treatment 1#Female				0.046
				(0.053)
Treatment 2# Female				0.062
				(0.051)
Treatment 3# Female				0.052
				(0.052)
Constant	0.565***	0.583***	0.588***	0.617***
	(0.034)	(0.035)	(0.037)	(0.040)
Observations	780	780	780	780
R-squared	0.027	0.026	0.023	0.021

Notes: Table explores heterogeneous treatment effects in true positives across demographic characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 20: Heterogeneous effects in true negatives: experience and behaviour

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.011	-0.012	0.011	0.031	-0.006	0.015
	(0.030)	(0.032)	(0.029)	(0.039)	(0.035)	(0.035)
Treatment 2	0.052*	0.002	0.048*	0.027	0.045	0.036
	(0.027)	(0.033)	(0.028)	(0.034)	(0.032)	(0.031)
Treatment 3	0.067**	0.002	0.045	0.036	0.023	0.035
	(0.028)	(0.032)	(0.028)	(0.036)	(0.033)	(0.031)
No direct experience	0.041					
	(0.037)					
Treatment 1# No direct experience	-0.026					
	(0.052)					
Treatment 2# No direct experience	-0.087*					
	(0.051)					
Treatment 3# No direct experience	-0.114**					
	(0.050)					
Low self-control		-0.007				
		(0.035)				
Treatment 1# Low self-control		0.023				
		(0.049)				
Treatment 2# Low self-control		0.039				
		(0.045)				
Treatment 3# Low self-control		0.059				
		(0.046)				
Low risk aversion			0.040			
			(0.037)			
Treatment 1#Low risk aversion			-0.027			
			(0.054)			
Treatment 2# Low risk aversion			-0.074			
			(0.047)			
Treatment 3# Low risk aversion			-0.038			
			(0.050)			
Low trust				0.058*		
				(0.035)		
Treatment 1# Low trust				-0.054		
				(0.049)		
Treatment 2# Low trust				-0.011		
				(0.045)		
Treatment 3# Low trust				-0.011		
				(0.047)		
Low DFS experience					-0.045	
					(0.035)	
Treatment 1# Low DFS experience					0.013	
					(0.049)	
Treatment 2# Low DFS experience					-0.047	
					(0.045)	
Treatment 3# Low DFS experience					0.019	
					(0.046)	
Low ICT experience						-0.010
						(0.035)
Treatment 1# Low ICT experience						-0.024
						(0.048)
Treatment 2# Low ICT experience						-0.027
						(0.045)
Treatment 3# Low ICT experience						-0.004
						(0.047)
Constant	0.699***	0.734***	0.706***	0.709***	0.716***	0.711***
	(0.032)	(0.033)	(0.032)	(0.036)	(0.034)	(0.034)
Observations	780	780	780	780	780	780
R-squared	0.070	0.064	0.065	0.064	0.065	0.062

Notes: Table explores heterogeneous treatment effects in accuracy across experiential and behavioural characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 21: Heterogeneous effects in true negatives: demographic

	(1)	(2)	(3)	(4)
Treatment 1	-0.019	0.014	0.015	0.014
	(0.032)	(0.032)	(0.032)	(0.037)
Treatment 2	-0.004	-0.015	0.033	0.039
	(0.029)	(0.030)	(0.031)	(0.035)
Treatment 3	-0.000	0.014	0.060**	0.060
	(0.029)	(0.031)	(0.029)	(0.037)
Agriculture	-0.072**			
	(0.036)			
Treatment 1#Agriculture	0.066			
	(0.049)			
Treatment 2#Agriculture	0.081*			
	(0.046)			
Treatment 3# Agriculture	0.097**			
	(0.047)			
Above median age		-0.016		
		(0.035)		
Treatment 1# Above median age		-0.029		
		(0.049)		
Treatment 2# Above median age		0.089**		
		(0.045)		
Treatment 3# Above median age		0.037		
		(0.046)		
Lower education			-0.051	
			(0.035)	
Treatment 1# Lower education			-0.020	
			(0.047)	
Treatment 2# Lower education			-0.016	
			(0.044)	
Treatment 3# Lower education			-0.049	
			(0.046)	
Female				-0.034
				(0.038)
Treatment 1#Female				-0.019
				(0.051)
Treatment 2# Female				-0.024
				(0.048)
Treatment 3# Female				-0.052
				(0.048)
Constant	0.738***	0.725***	0.745***	0.701***
	(0.032)	(0.032)	(0.034)	(0.037)
Observations	780	780	780	780
R-squared	0.067	0.071	0.083	0.063

Notes: Table explores heterogeneous treatment effects in true negatives across demographic characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 22: Heterogeneous effects in confidence: experience and behaviour

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	-0.019	0.060	-0.061	0.211	-0.099	0.167
	(0.135)	(0.133)	(0.134)	(0.189)	(0.152)	(0.161)
Treatment 2	0.186	0.210	0.158	0.232	0.120	0.372**
	(0.129)	(0.133)	(0.127)	(0.186)	(0.154)	(0.161)
Treatment 3	0.148	0.059	0.162	0.499***	0.029	0.440***
	(0.137)	(0.150)	(0.134)	(0.183)	(0.154)	(0.163)
No direct experience	-0.219					
	(0.179)					
Treatment 1# No direct experience	0.069					
	(0.254)					
Treatment 2# No direct experience	0.035					
	(0.257)					
Treatment 3# No direct experience	0.368					
	(0.256)					
Low self-control		-0.343**				
		(0.167)				
Treatment 1# Low self-control		-0.161				
		(0.243)				
Treatment 2# Low self-control		0.002				
		(0.212)				
Treatment 3# Low self-control		0.403*				
		(0.225)				
Low risk aversion			-0.600***			
			(0.189)			
Treatment 1#Low risk aversion			0.209			
			(0.256)			
Treatment 2# Low risk aversion			0.131			
			(0.246)			
Treatment 3# Low risk aversion			0.296			
			(0.255)			
Low trust				0.638***		
				(0.168)		
Treatment 1# Low trust				-0.384*		
				(0.231)		
Treatment 2# Low trust				-0.070		
				(0.221)		
Treatment 3# Low trust				-0.477**		
				(0.227)		
Low DFS experience					-0.416**	
					(0.171)	
Treatment 1# Low DFS experience					0.204	
					(0.233)	
Treatment 2# Low DFS experience					0.163	
					(0.219)	
Treatment 3# Low DFS experience					0.478**	
					(0.231)	
Low ICT experience						0.075
						(0.171)
Treatment 1# Low ICT experience						-0.292
						(0.224)
Treatment 2# Low ICT experience						-0.326
						(0.220)
Treatment 3# Low ICT experience						-0.359
						(0.232)
Constant	5.939***	5.936***	5.961***	5.788***	6.004***	5.758***
	(0.155)	(0.146)	(0.153)	(0.175)	(0.159)	(0.170)
Observations	780	780	780	780	780	780
R-squared	0.167	0.172	0.166	0.171	0.169	0.168

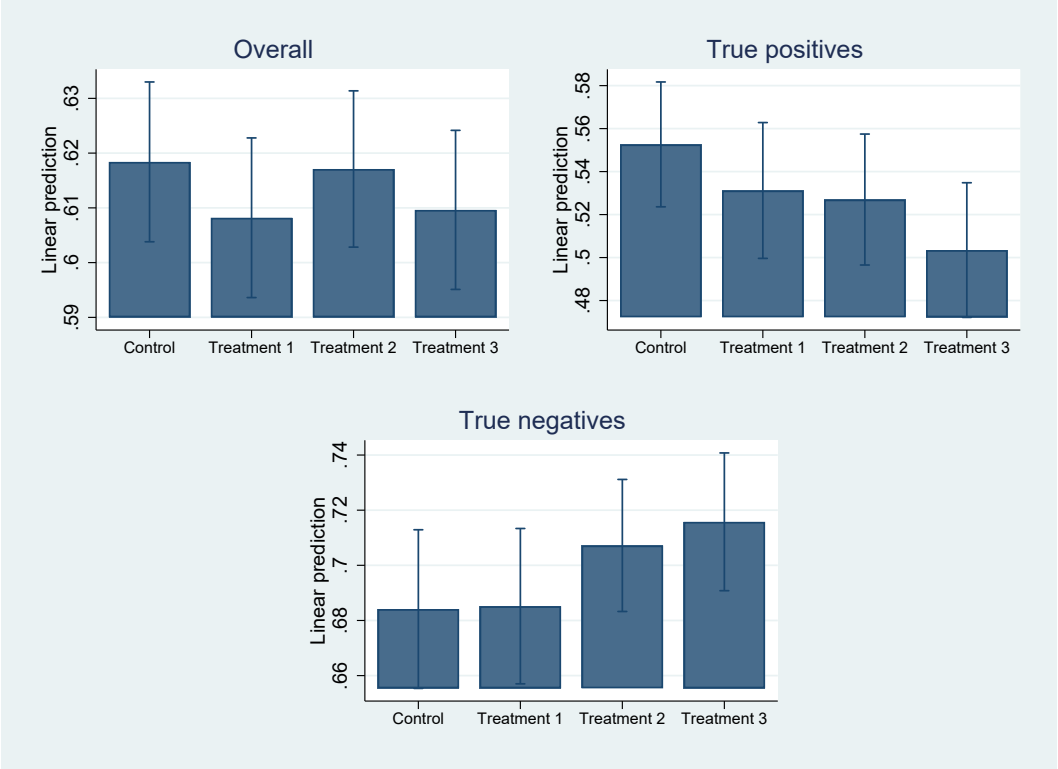
Notes: Table explores heterogeneous treatment effects in confidence across experiential and behavioural characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 23: Heterogeneous effects in confidence: demographic

	(1)	(2)	(3)	(4)
Treatment 1	-0.018	0.028	-0.043	-0.211
	(0.146)	(0.150)	(0.153)	(0.167)
Treatment 2	0.210	0.109	0.166	0.090
	(0.145)	(0.153)	(0.148)	(0.156)
Treatment 3	0.396***	0.403***	0.176	-0.035
	(0.139)	(0.156)	(0.144)	(0.169)
Agriculture	0.090			
	(0.177)			
Treatment 1#Agriculture	0.043			
	(0.239)			
Treatment 2#Agriculture	-0.035			
	(0.221)			
Treatment 3# Agriculture	-0.390			
	(0.239)			
Above median age		0.011		
		(0.171)		
Treatment 1# Above median age		-0.041		
		(0.235)		
Treatment 2# Above median age		0.237		
		(0.219)		
Treatment 3# Above median age		-0.260		
		(0.229)		
Lower education			-0.233	
			(0.168)	
Treatment 1# Lower education			0.092	
			(0.225)	
Treatment 2# Lower education			0.071	
			(0.216)	
Treatment 3# Lower education			0.149	
			(0.225)	
Female				-0.497***
				(0.173)
Treatment 1#Female				0.380
				(0.232)
Treatment 2# Female				0.137
				(0.227)
Treatment 3# Female				0.528**
				(0.231)
Constant	5.872***	5.882***	6.037***	6.084***
	(0.160)	(0.150)	(0.159)	(0.171)
Observations	780	780	780	780
R-squared	0.169	0.169	0.168	0.171

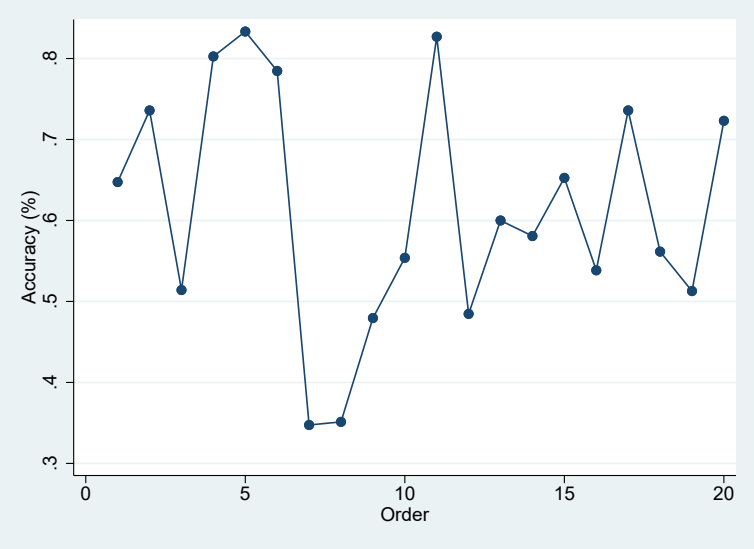
Notes: Table explores heterogeneous treatment effects in confidence across demographic characteristics, using interaction terms. Regression includes vector of controls listed in Table 11. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Direction of treatment effects in overall accuracy



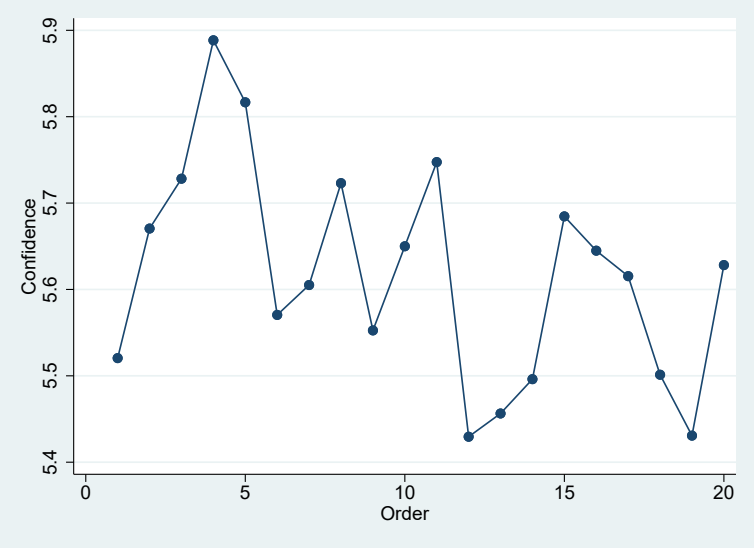
Note: Figure reports marginal effects corresponding to Table 3, depicting the direction of treatment effects in accuracy outcomes. While we do not see significant treatment effects, directional patterns are evident.

Figure 2: Accuracy in judgements through the sequence of scenarios



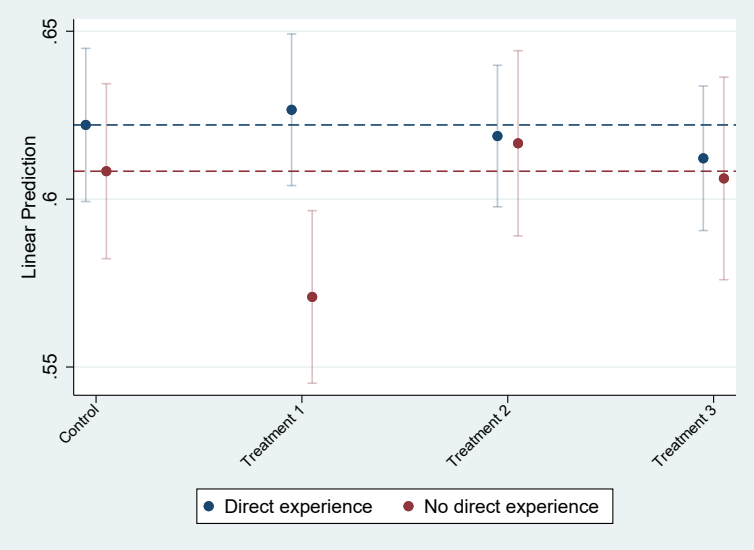
Note: Figure plots the mean level of accuracy in scenario judgements by the order in which they appear in the experimental task. The placement of any individual scenario in the sequence is randomised. Figure graphically evaluates the hypothesis that performance shows a 'learning by doing' effect.

Figure 3: Level of confidence reported through the sequence of scenarios



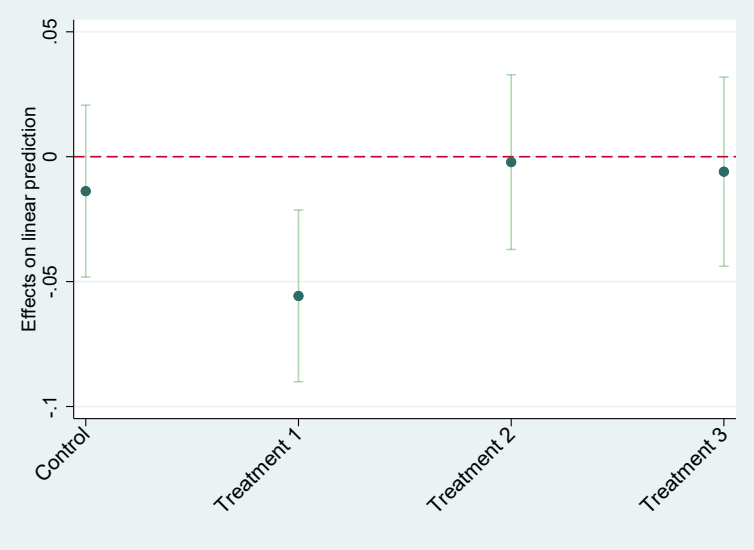
Note: Figure plots the mean level of confidence reported in scenario judgements by the order in which they appear in the experimental task. The placement of any individual scenario in the sequence is randomised.

Figure 4: Predictive margins in accuracy (fraud experience interaction)



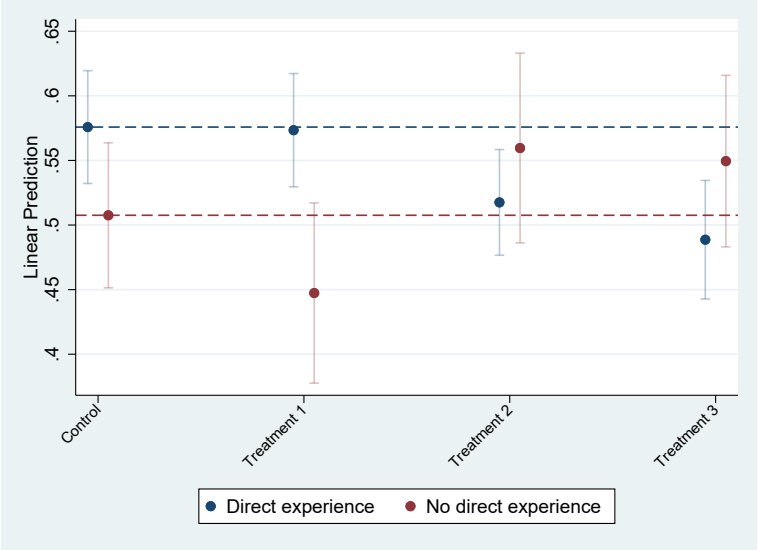
Note: Figure reports predicted outcomes in accuracy, from an interaction of treatment status with fraud experience found in Table 16. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 5: Difference in predicted values (no direct experience:1, direct experience:0)



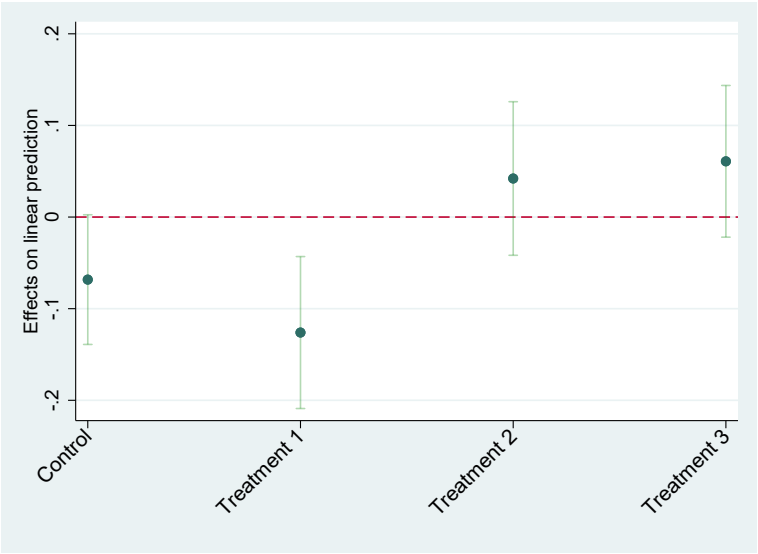
Note: Figure plots the relationship shown in Figure 4, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 6: Predictive margins in true positives (fraud experience interaction)



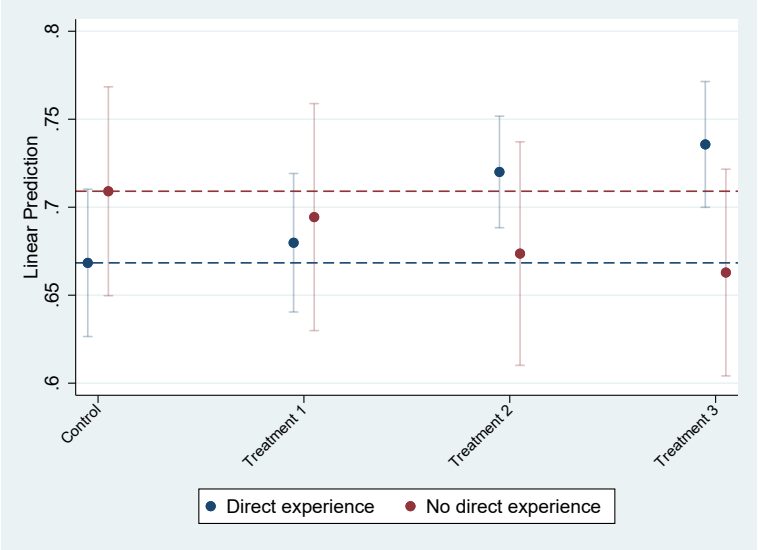
Note: Figure reports predicted outcomes in true positives, from an interaction of treatment status with fraud experience found in Table 18. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 7: Difference in predicted values (no direct experience:1, direct experience:0)



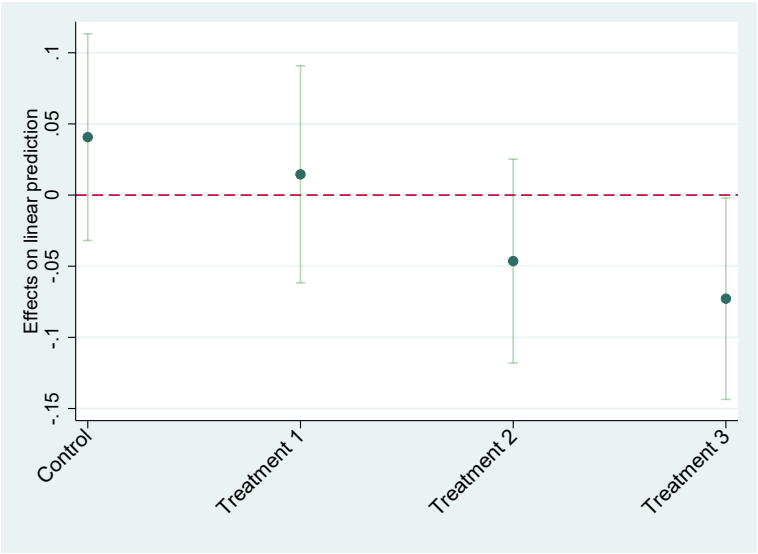
Note: Figure plots the relationship shown in Figure 6, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 8: Predictive margins in true negatives (fraud experience interaction)



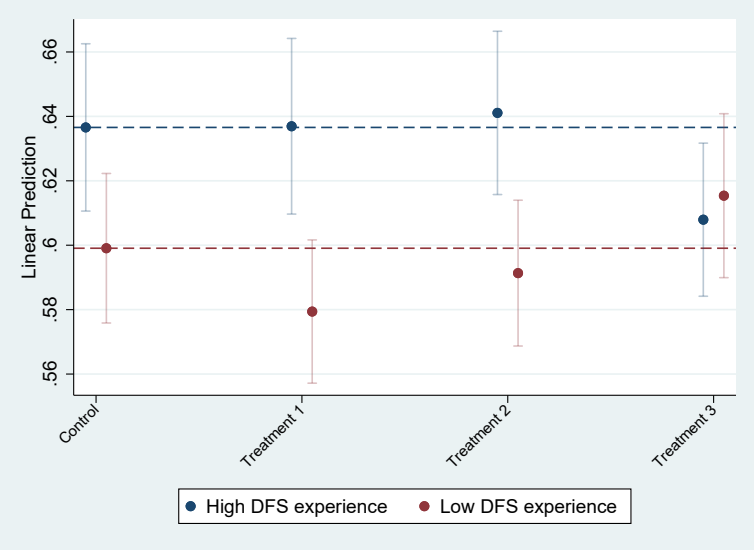
Note: Figure reports predicted outcomes in true negatives, from an interaction of treatment status with fraud experience found in Table 20. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 9: Difference in predicted values (no direct experience:1, direct experience:0)



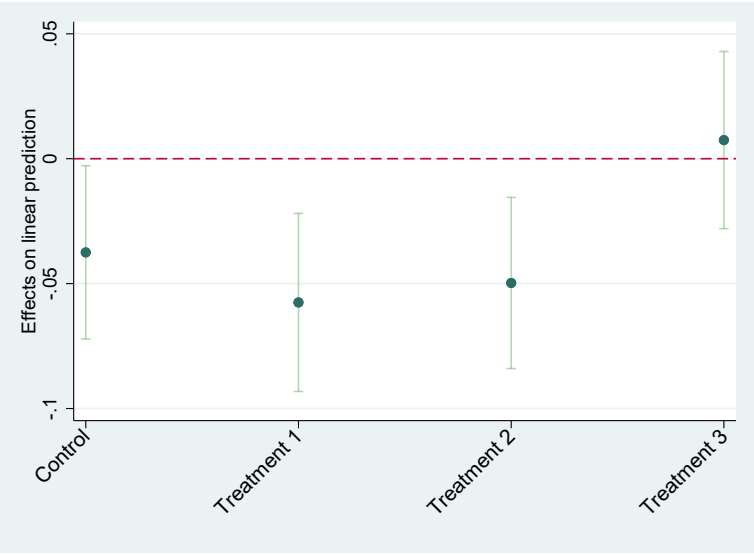
Note: Figure plots the relationship shown in Figure 8, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 10: Predictive margins in accuracy (DFS experience interaction)



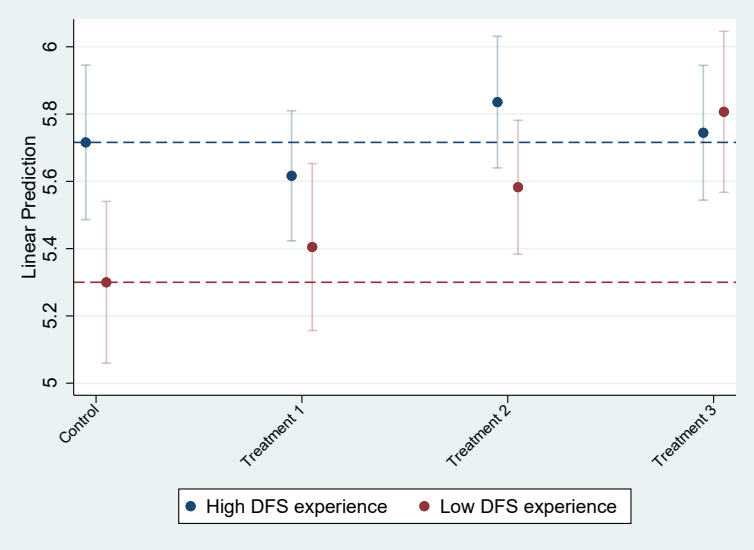
Note: Figure reports predicted outcomes in accuracy, from an interaction of treatment status with DFS experience found in Table 20. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with level difference evident.

Figure 11: Difference in predicted values (Low DFS experience:1, High DFS experience:0)



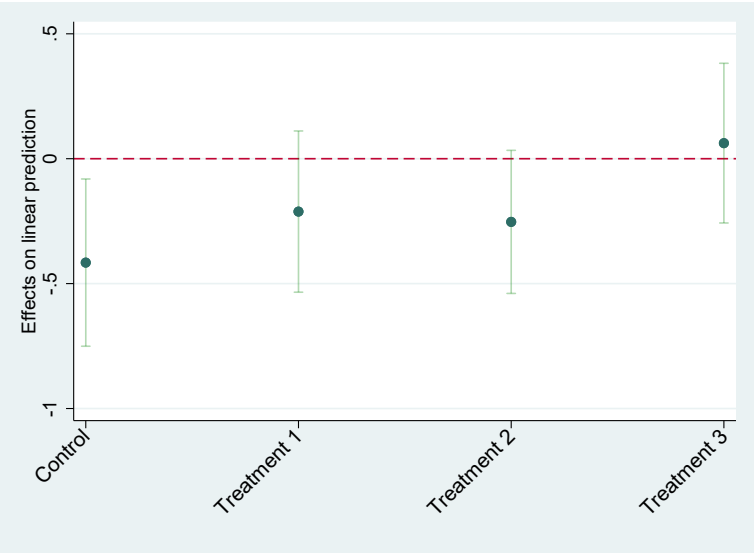
Note: Figure plots the relationship shown in Figure 10, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 12: Predictive margins in confidence (DFS experience interaction)



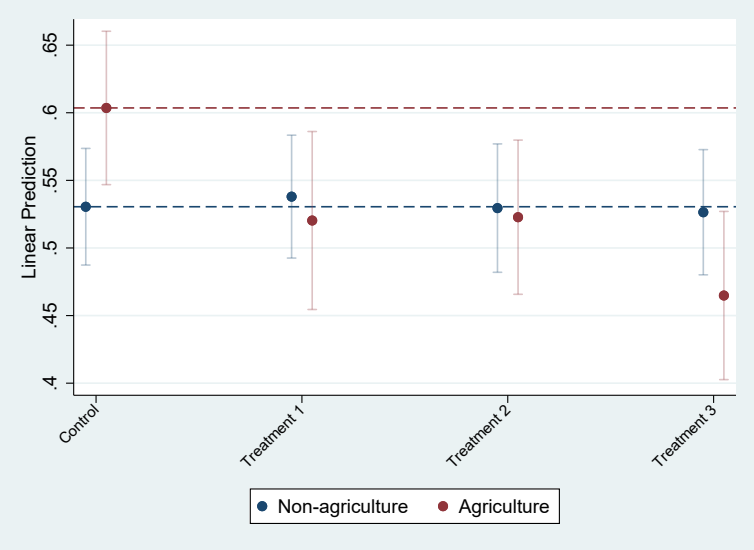
Note: Figure reports predicted outcomes in confidence, from an interaction of treatment status with DFS experience found in Table 22. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 13: Difference in predicted values (Low DFS experience:1, High DFS experience:0)



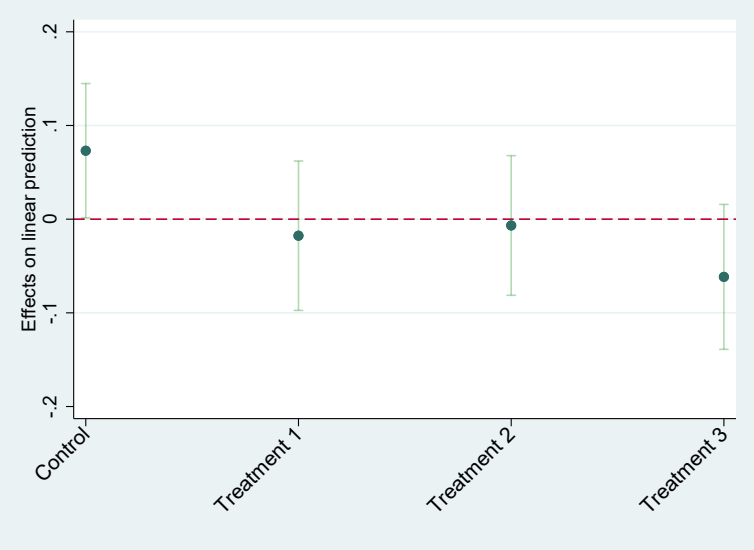
Note: Figure plots the relationship shown in Figure 12, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 14: Predictive margins in true positives (employment sector interaction)



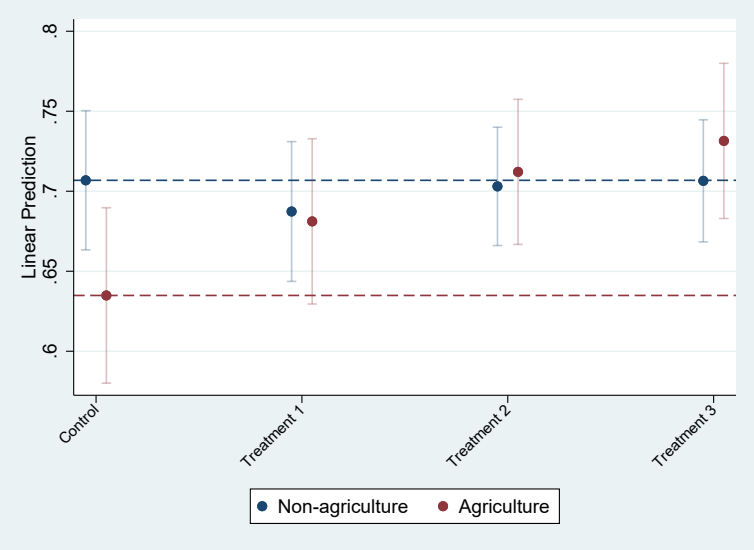
Note: Figure reports predicted outcomes in true positives, from an interaction of treatment status with employment sector found in Table 19. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 15: Difference in predicted values (Agriculture:1, Non-agriculture:0)



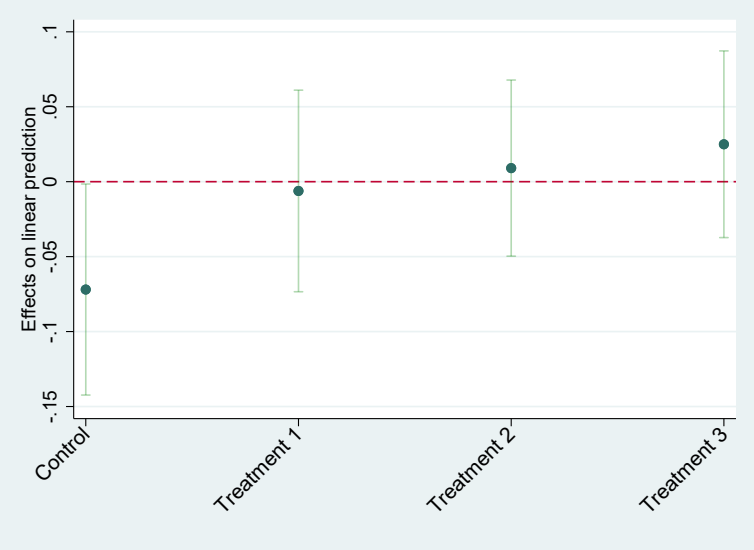
Note: Figure plots the relationship shown in Figure 14, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 16: Predictive margins in true negatives (employment sector interaction)



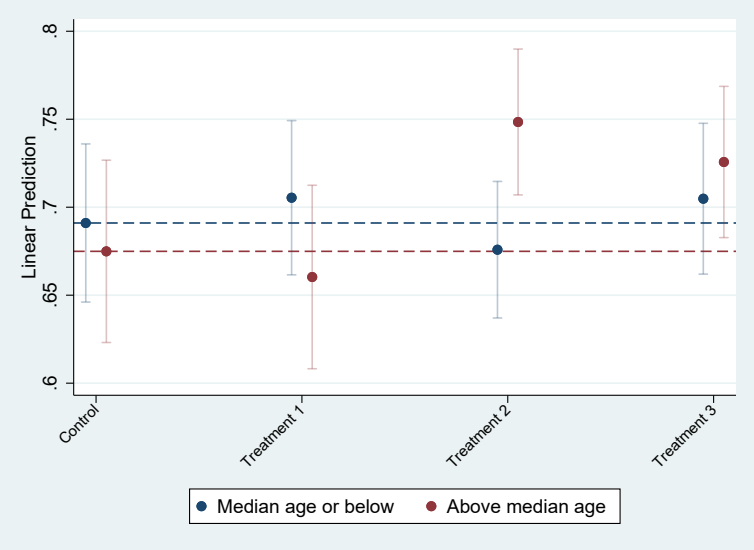
Note: Figure reports predicted outcomes in true negatives, from an interaction of treatment status with employment sector found in Table 21. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 17: Difference in predicted values (Agriculture:1, Non-agriculture:0)



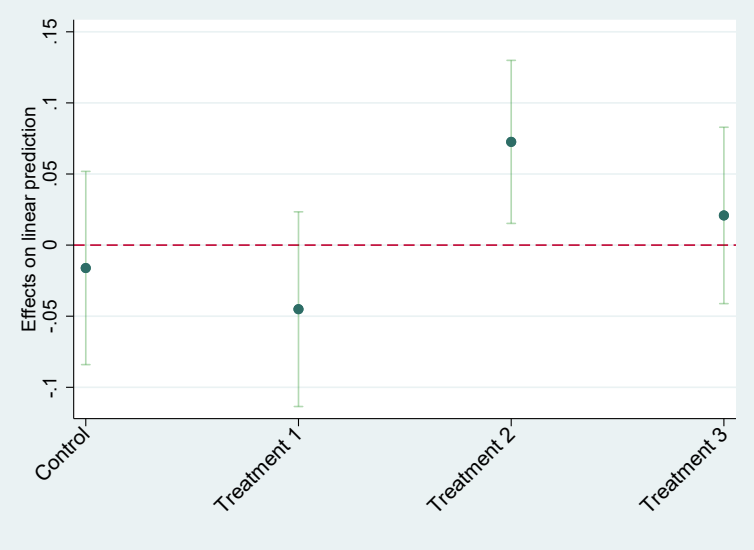
Note: Figure plots the relationship shown in Figure 16, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 18: Predictive margins in true negatives (age interaction)



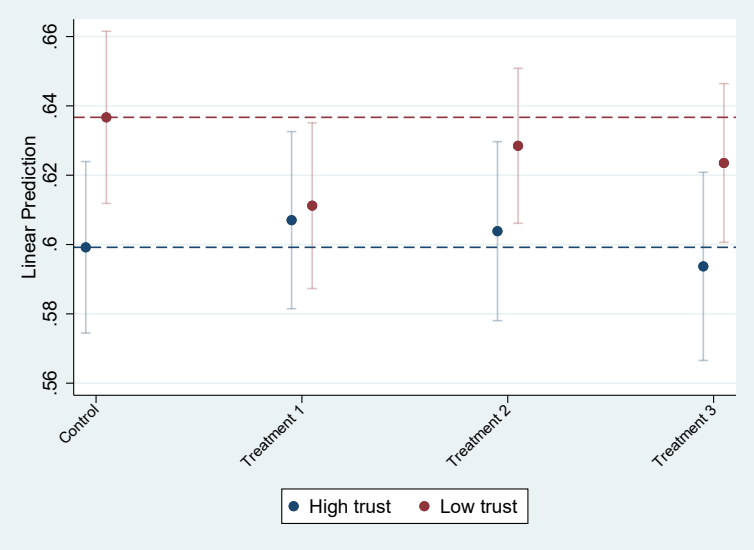
Note: Figure reports predicted outcomes in true negatives, from an interaction of treatment status with age found in Table 21. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 19: Difference in predicted values (Above median age:1, Median age or below:0)



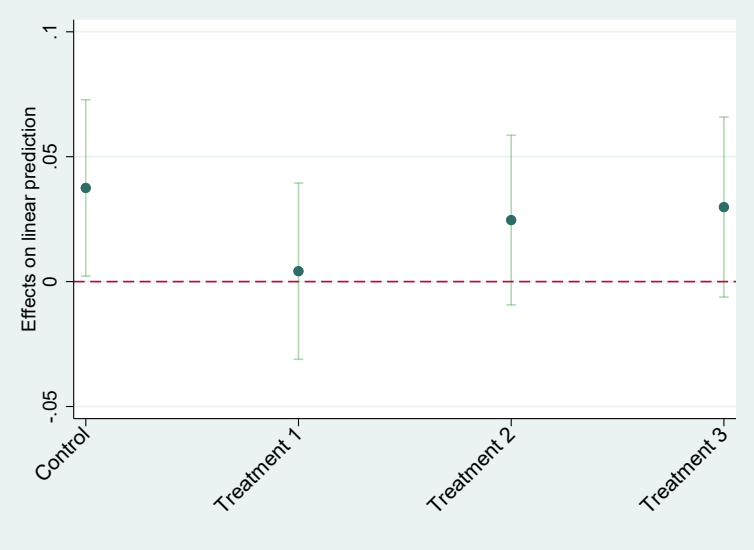
Note: Figure plots the relationship shown in Figure 18, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 20: Predictive margins in accuracy (trust interaction)



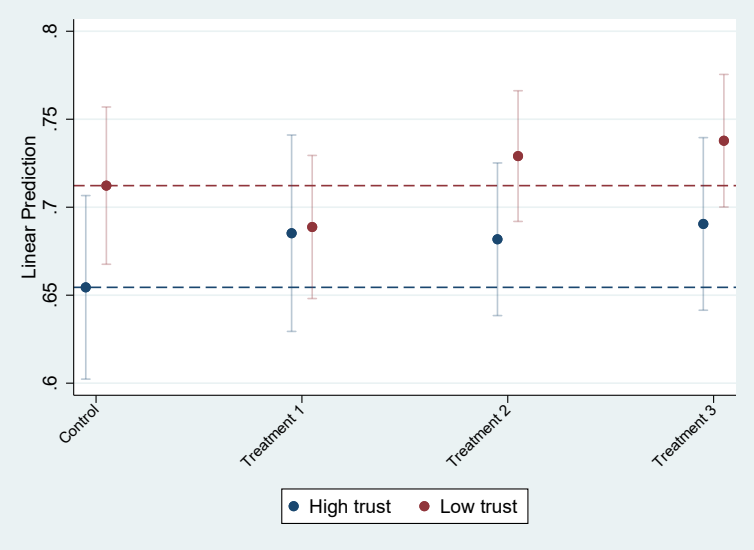
Note: Figure reports predicted outcomes in accuracy, from an interaction of treatment status with trust found in Table 16. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with tentative level difference evident.

Figure 21: Difference in predicted values (Low trust:1, High trust:0)



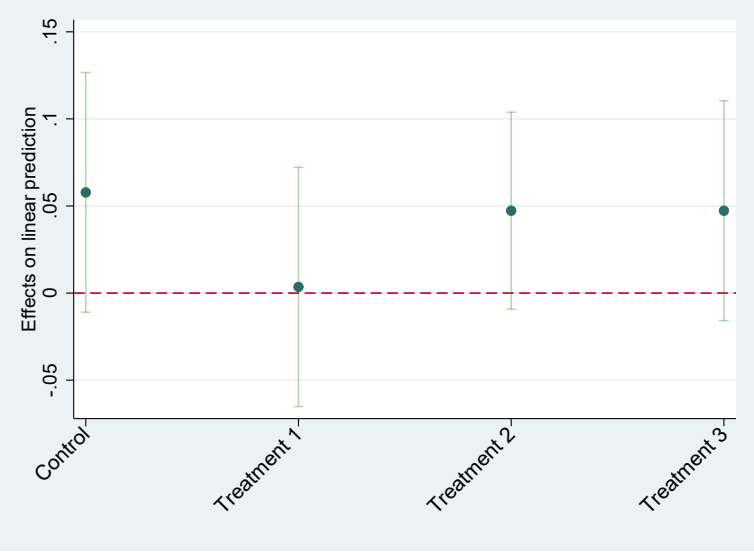
Note: Figure plots the relationship shown in Figure 20, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 22: Predictive margins in true negatives (trust interaction)



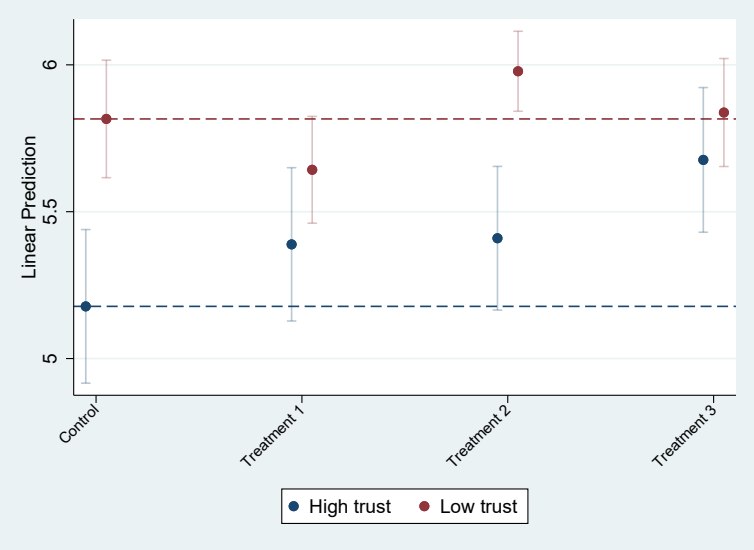
Note: Figure reports predicted outcomes in true negatives, from an interaction of treatment status with trust found in Table 20. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation.

Figure 23: Difference in predicted values (Low trust:1, High trust:0)



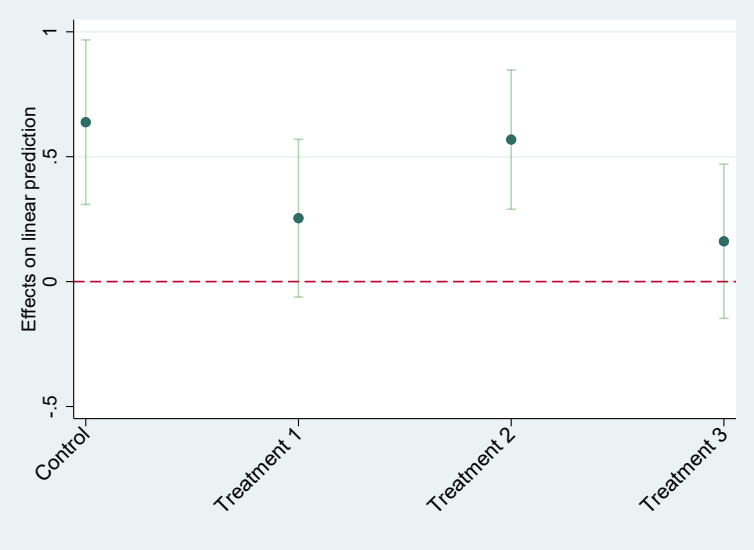
Note: Figure plots the relationship shown in Figure 22, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 24: Predictive margins in confidence (trust interaction)



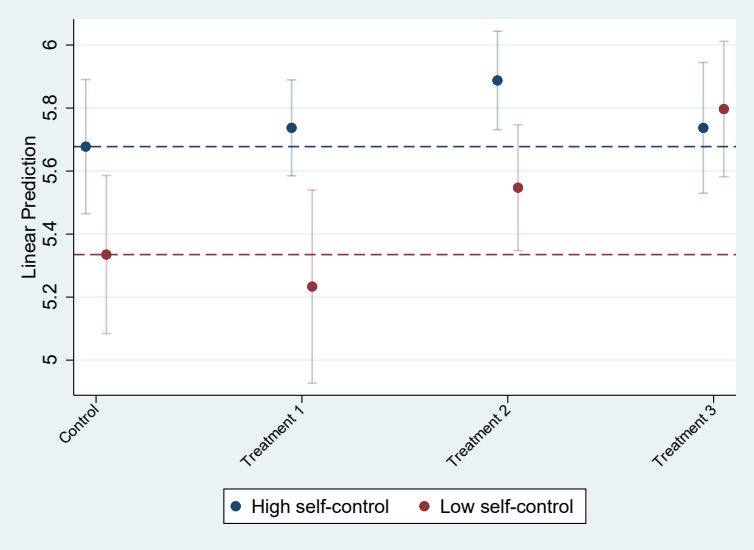
Note: Figure reports predicted outcomes in confidence, from an interaction of treatment status with trust found in Table 22. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with level difference evident.

Figure 25: Difference in predicted values (Low trust:1, High trust:0)



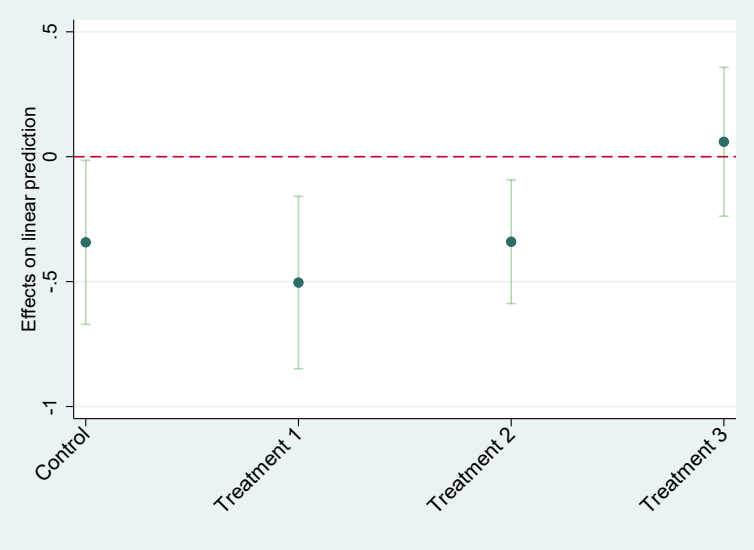
Note: Figure plots the relationship shown in Figure 24, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 26: Predictive margins in confidence (self-control interaction)



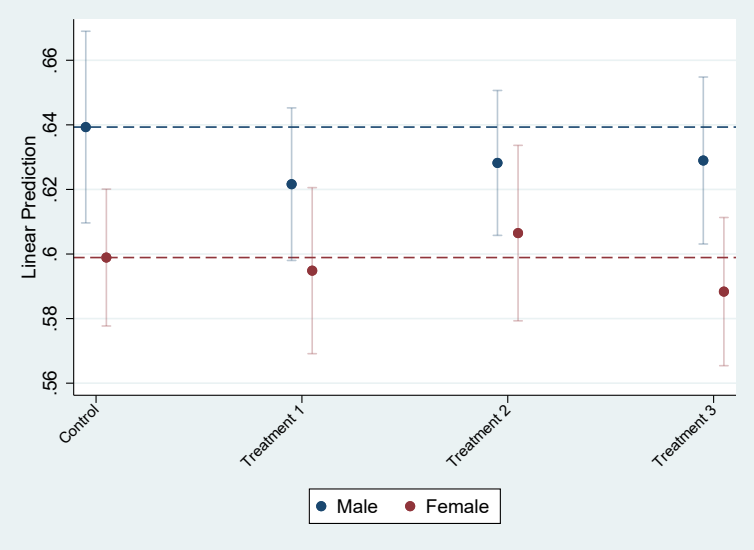
Note: Figure reports predicted outcomes in confidence, from an interaction of treatment status with trust found in Table 22. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with level difference evident.

Figure 27: Difference in predicted values (Low self-control:1, High self-control:0)



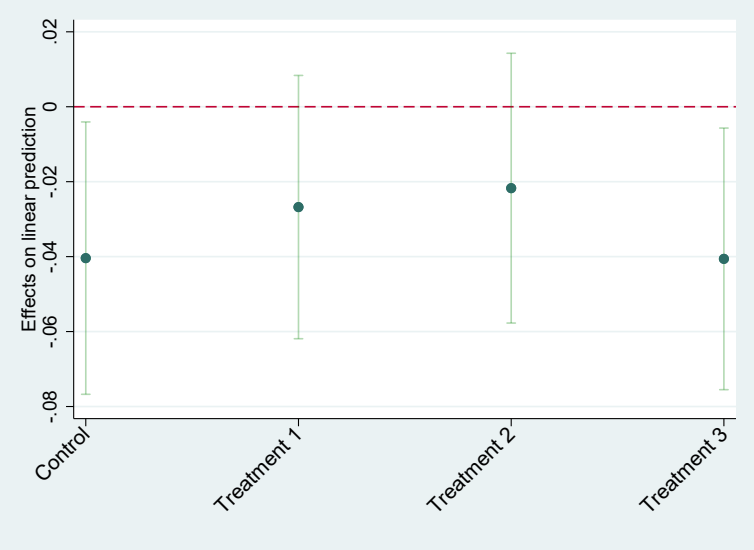
Note: Figure plots the relationship shown in Figure 26, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 28: Predictive margins in accuracy (gender interaction)



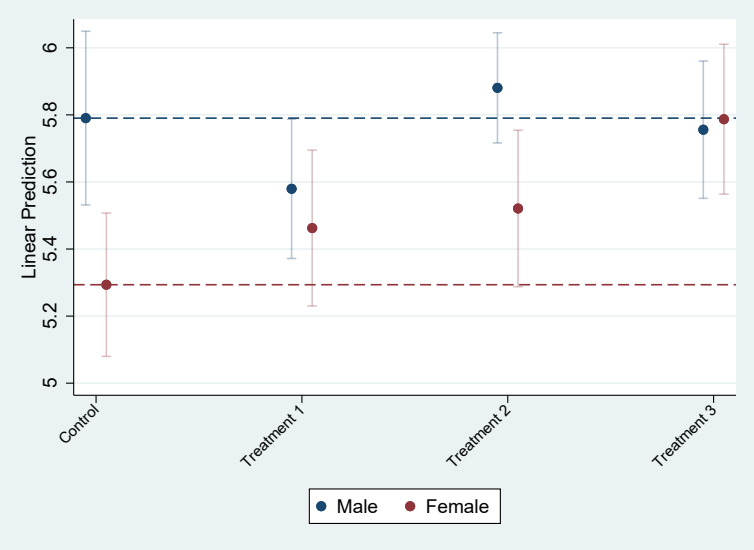
Note: Figure reports predicted outcomes in accuracy, from an interaction of treatment status with gender found in Table 17. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with tentative level difference evident.

Figure 29: Difference in predicted values (Female:1, Male:0)



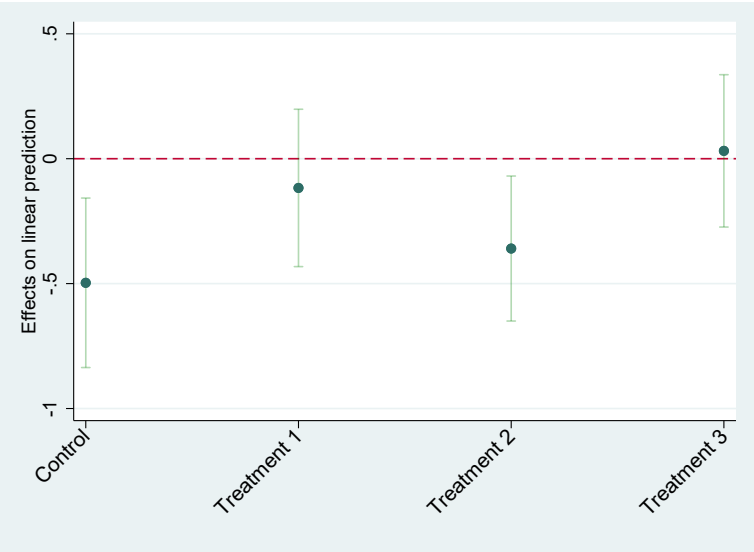
Note: Figure plots the relationship shown in Figure 28, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 30: Predictive margins in confidence (gender interaction)



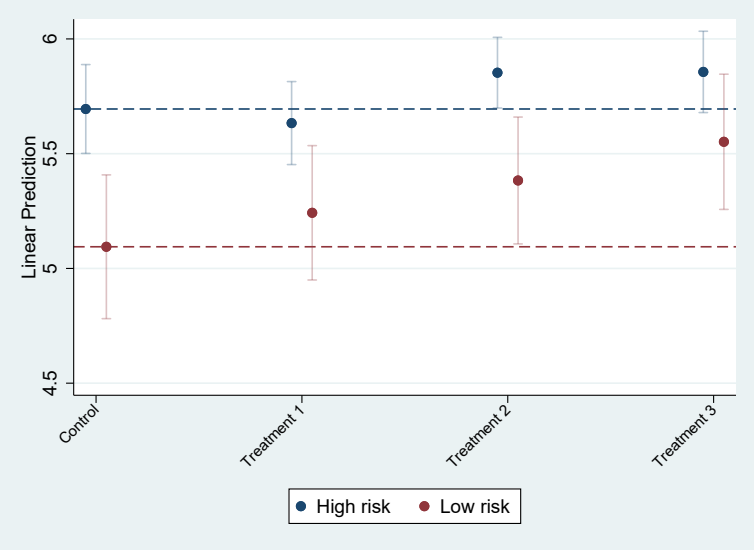
Note: Figure reports predicted outcomes in confidence, from an interaction of treatment status with gender found in Table 23. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with tentative level difference evident.

Figure 31: Difference in predicted values (Female:1, Male:0)



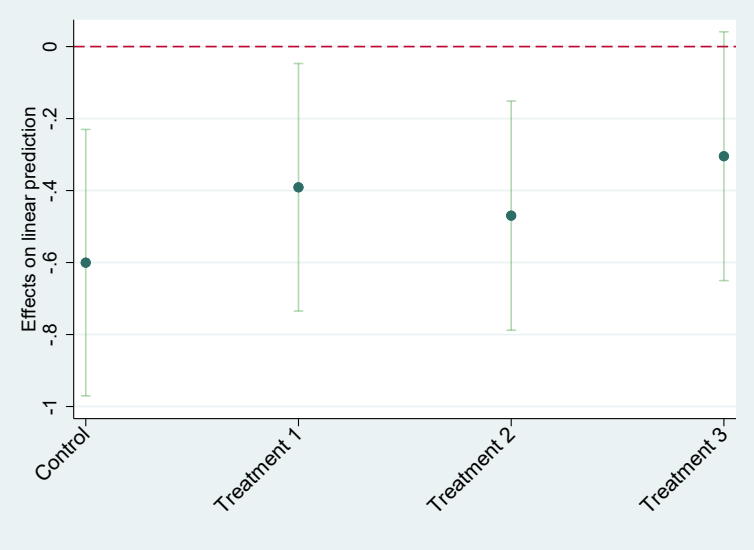
Note: Figure plots the relationship shown in Figure 30, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Figure 32: Predictive margins in confidence (risk appetite interaction)



Note: Figure reports predicted outcomes in confidence, from an interaction of treatment status with risk appetite found in Table 22. Superficial appearance of significant heterogeneity in treatment effect found in the regression table is shown to be immaterial on graphical representation, albeit with level difference evident.

Figure 33: Difference in predicted values (Low risk appetite:1, High risk appetite:0)



Note: Figure plots the relationship shown in Figure 32, but focuses on the difference across the moderating variable at each treatment cell, which has its own error term. This is to establish whether that difference is statistically different to zero.

Table 24: Overall effect - pooled treatments

	(1) Overall	(2) Overall	(3) True positive	(4) True positive	(5) True negative	(6) True negative
Treated	-0.007		-0.032		0.019	
	(0.010)		(0.021)		(0.020)	
Treated (2 or 3)		-0.005		-0.036		0.026
		(0.011)		(0.022)		(0.021)
Constant	0.654***	0.642***	0.591***	0.582***	0.718***	0.702***
	(0.016)	(0.018)	(0.033)	(0.035)	(0.030)	(0.033)
Observations	780	585	780	585	780	585
R-squared	0.084	0.077	0.018	0.018	0.059	0.073
p-value ($\beta \leq 0$)	0.746	0.671	0.939	0.945	0.174	0.103

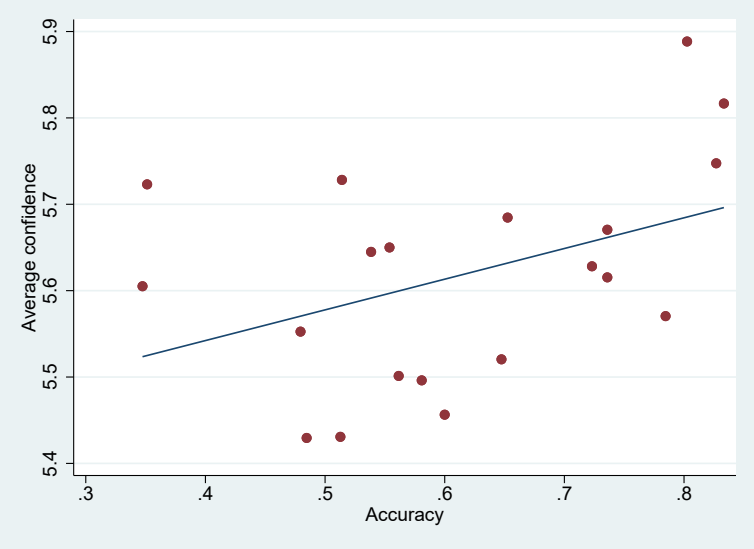
Notes: Table reports pooled treatment effects in overall accuracy, true positives, and true negatives, as an aggregated counterpart to Table 3. Columns 1, 3, and 5 report all treatment arms pooled against the control group, while Columns 2, 4, and 6 pool only Treatments 2 and 3 against the control group. Table reports results from two-sided test for pooled treatment effects on overall accuracy, true positives, and true negatives. Also reported are one-sided tests of pre-specified hypotheses for incremental positive treatment effects from each treatment arm compared against the preceding arm in the sequence. Regression includes vector of controls listed in Table 11. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 25: Confidence effect - pooled treatments

	(1) Overall	(2) Overall	(3) Genuine	(4) Genuine	(5) Fraudulent	(6) Fraudulent
Treated	0.157		0.134		0.173*	
	(0.096)		(0.098)		(0.100)	
Treated (2 or 3)		0.222**		0.193*		0.241**
		(0.102)		(0.105)		(0.105)
Constant	5.896***	5.848***	5.853***	5.785***	5.930***	5.893***
	(0.144)	(0.154)	(0.147)	(0.158)	(0.148)	(0.161)
Observations	780	585	780	585	780	585
R-squared	0.158	0.175	0.146	0.159	0.153	0.173
p-value ($\beta \leq 0$)	0.0512	0.0152	0.0864	0.0332	0.0416	0.0109

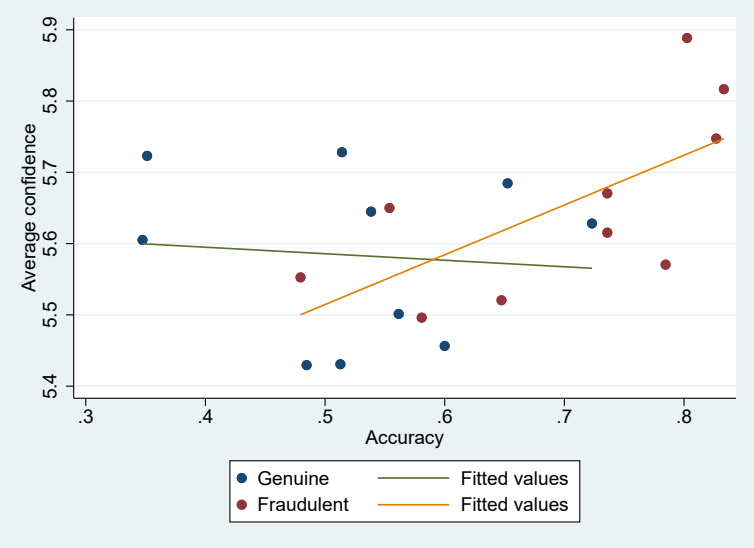
Notes: Table reports pooled treatment effects in overall confidence, confidence in reported in genuine calls, and confidence reported in fraudulent calls, as an aggregated counterpart to Table 4. Columns 1, 3, and 5 report all treatment arms pooled against the control group, while Columns 2, 4, and 6 pool only Treatments 2 and 3 against the control group. Table reports results from two-sided test for pooled treatment effects on overall accuracy, true positives, and true negatives. Also reported are one-sided tests of pre-specified hypotheses for incremental positive treatment effects from each treatment arm compared against the preceding arm in the sequence. Regression includes vector of controls listed in Table 11. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 34: Accuracy and confidence: scenario averages



Note: Figure describes a scatter plot of mean accuracy achieved in each scenario against mean confidence reported by participants in their judgement on that scenario, and shows a positive correlation.

Figure 35: Accuracy and confidence: scenario averages, separated by status



Note: Figure describes a scatter plot of mean accuracy achieved in each scenario against mean confidence reported by participants in their judgement on that scenario, separating by genuine and fraudulent scenarios. The positive overall correlation observed in Figure 34 is shown to be driven by true negatives.

Table 26: Key signs of fraud

Key sign	Description
Fabricated sense of urgency	Pressuring you to "act now" or else a deal will go away, your account will be closed, or you will experience other negative consequences.
Random outreach	You are contacted out of the blue, e.g., the message comes from an unfamiliar email address, behind what looks like a genuine sender name, or phone call etc. and it is hard to understand why you are being contacted.
Unfamiliar but genuine looking email	The message comes from an unfamiliar email address, behind what looks like a genuine sender name.
Poorly written message	The message is poorly written with misspellings and incorrect grammar, or a familiar company name is misspelt.
Personal information theft	Asking for personal information and access to your money-such as your ATM cards, bank accounts, credit cards, or investment account, or for you to confirm personal information they claim to have.
Suspicious call	Calling or emailing you, claiming to be from the government and asking you to pay money.
Suspicious offer	The offer seems too good to be true.

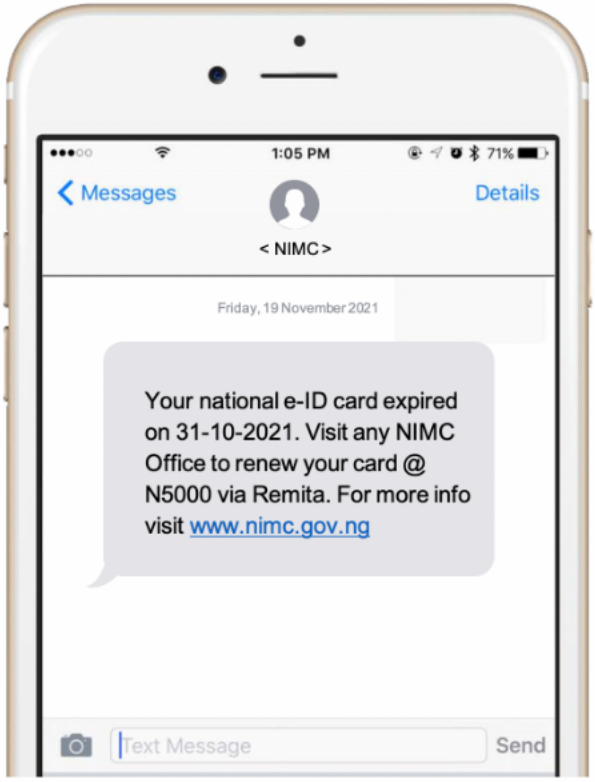
Notes: Table lists the seven key signs of fraud included in Treatments 2 and 3.

Table 27: Overview of scenarios

Scenario	Genuine/Fraudulent	Description
1	F	Bank email account update
2	F	Pop-up window link to claim prize
3	G	Bank text account update
4	F	Mobile company call for sensitive info
5	F	Call re investment opportunity, seeks transfer and personal info
6	F	WhatsApp lawyer inheritance, seeks processing fee and personal info
7	G	SMS chance to win
8	G	SMS gov ID expired
9	F	Email delivery update fee
10	F	Call+SMS re. accidental cash transfer
11	F	Call re. prize giveaway, seeks processing fee
12	G	Email annual account statement
13	G	Email survey
14	F	Email investment opportunity
15	G	Email order collection
16	G	Call re. gov loan scheme
17	F	WhasApp offer with fee
18	G	Email reward offer
19	G	Email cash offer
20	G	Email customer survey

Notes: Table reports an overview of the 20 scenarios which participants are asked to evaluate as part of the experimental task.

Figure 36: Example of a genuine scenario in the experimental task

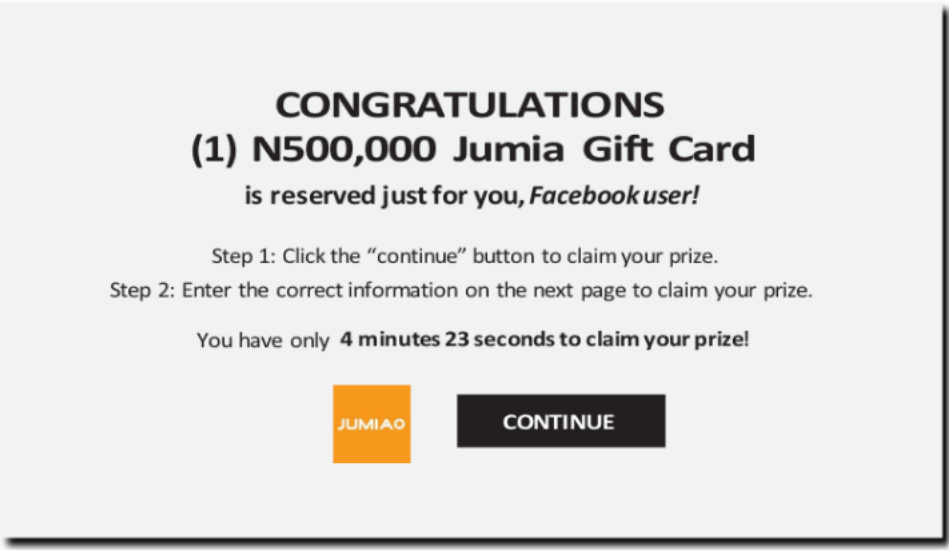


Do you think this is a genuine or fraudulent scenario?

Kana/kina ganin wannan labari ba na damfara bane ko kuwa na damfara ne?

- ◯ Genuine
  Ba damfara bace
- ◯ Fraudulent
  Damfara ce

Figure 37: Example of a fraudulent scenario in the experimental task



Do you think this is a genuine or fraudulent scenario?
Kana/kina ganin wannan labari ba na damfara bane ko kuwa na damfara ne?

- Genuine
  Ba damfara bace
- Fraudulent
  Damfara ce

Figure 38: Elicitation of confidence in judgements following each presented scenario

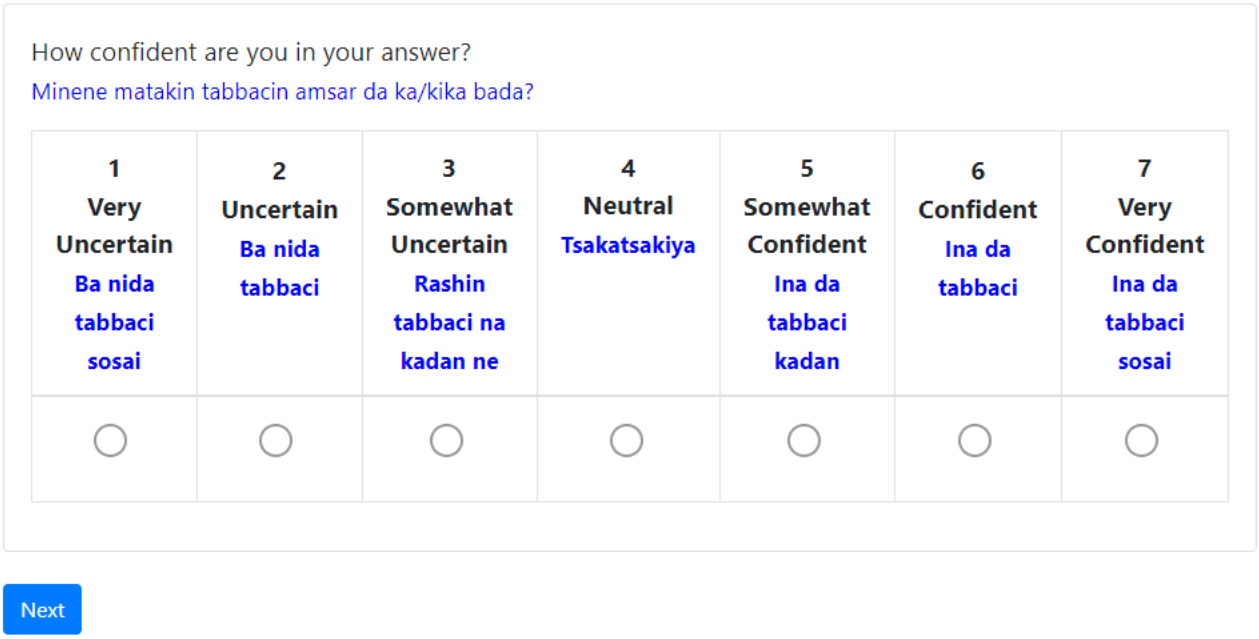


Table 28: Definition of variables used

Variable	Definition
No direct experience	Binary variable recording that the participant reports 'No' to the question: • Have you ever experienced someone contacting you pretending to be someone else to steal money or sensitive information?
Low self-control	Binary variable recording that the participant falls below the median value of a standardised index constructed from the participantâs level of agreement with the following questions: • I spend too much in the moment and let the future take care of itself • Financial services are complicated and confusing to me • Convenience plays an important role in the decisions I make • I often act without thinking through all the alternatives • I am optimistic about my future • If I work hard today, I will be more successful in the future

Variable	Definition
Low risk appetite	Binary variable recording that the participant falls below the median value of a standardised index constructed from the participant's responses to the following questions: • Suppose you're offered a business investment that returns 5,000 Naira on average. Half the time the investment returns 10,000 Naira. However, half of the time the investment returns nothing. What is the maximum you personally would be willing to pay to make this investment? • Indicate your level of agreement for the following statement: I am a person who takes risks
Low trust	Binary variable recording that the participant falls below the median value of a standardised index constructed from the participant's responses to the following questions: • In general, most people can be trusted • I often reject statements unless I have proof that they are true • I frequently question things that I see or hear

Variable	Definition
Low DFS experience	Binary variable recording that the participant falls below the median value of a standardised index constructed from the participant's responses to the following questions: • Do you have an account at a formal financial institution? • Have you used a formal bank account in the last 90 days? • When did you first get a formal bank account? • Can you access your formal bank account or bank application on your phone? • Have you accessed your formal bank account on your phone in the last 90 days? • Do you use a phone for conducting business? • Have suppliers contacted you on your personal phone or business phone in the last 90 days? • Do you have a mobile money account? E.g. Paga Mobile, MTN Momo, First Banks Firstmonie, Kudi Mobile, UBA Moni Agent or Polaris Sure Padi • Have you ever transferred money to another individual or business using your phone? • Have you used mobile money or any other digital payments provider to send or receive payments in the last 90 days? • When did you first use mobile money or any other digital payment services to send or receive payments? • Do you usually access mobile money or other digital payment services using an app, shortcode menu or SMS? • Have you ever bought or sold goods using an online platform (e.g. Jumia)? • When was the first time you bought or sold goods using an online platform?

Variable	Definition
Low ICT experience	Binary variable recording that the participant falls below the median value of a standardised index constructed from the participant's responses to the following questions: • Do you have access to a smartphone? • Are you the primary user for your smartphone? • On average over the past 30 days, how often have you used a smartphone to do any of the following: – To make phone calls? – To send SMS messages? – Use messenger apps (Facebook messenger, WhatsApp, etc.) – Browse social media? – To conduct purchasing transactions – To conduct banking transactions
Agriculture	Binary variable recording whether the participant is employed in the agricultural sector (1), or a non-agricultural sector (0).
Above median age	Participant age is greater than 25 years.
Lower education	Binary variable recording whether the participantâs level of educational attainment is secondary or below (1), or tertiary (0).
Female	Binary variable recording whether the participant identifies as female (1), or not (0).

Variable	Definition
ICW trust index	Inverse correlation weighted index of trust constructed from the participant's responses to the following questions (computed at baseline and endline): How likely are you to use these types of DFSs (Digital Financial Services) in the future? • Banks • Mobile banking • Mobile money operators • Online platforms for buying or selling goods • Agents Indicate your level of agreement for the following statement: In general, I trust that my financial information is kept safe by - • Banks • Mobile banking • Mobile money operators • Online platforms for buying or selling goods • Agents Indicate your level of agreement for the following statement: In general, I trust that my money is kept safe from fraud or theft by: • Banks • Mobile banking • Mobile money operators • Online platforms for buying or selling goods • Agents
Age	Reported age of participant.
Third level education	Binary inverse of 'Lower education' variable.

Variable	Definition
Married	Binary variable recording whether the participant is married (1) or not married (i.e. divorced, separated, single, or widowed) (0).
Contacted by scammer	Binary inverse of âNo direct experienceâ variable.
Access to smartphone	Binary variable recording whether the participant has access to a smartphone (1), or not (0).
Business owner	Binary variable recording whether the participant is a business owner (1), or not (0).
Has formal financial account	Binary variable recording whether the participant answered has an account at a formal financial institution (1), or not (0).
Used online platforms	Binary variable recording whether the participant has ever bought or sold goods using an online platform (1), or not (0).
Trusting	Binary inverse of 'Low trust' variable.
Risk averse	See 'Low risk appetite'.

Notes: Table reports a description of each variable used in the analysis.