Strategies for Reducing Non-Institutional Fraud and Building Trust in a Digital Market Platform: A Behavioral Lab Experiment in Nigeria

Pre-Analysis Plan

Daniel Putman, Michael King, Shane Byrne, and Chaning Jang

Abstract: The high prevalence of digital financial fraud makes it difficult for businesses to distinguish between real communications from digital service providers and fraudulent communication. This could lead to a lack of trust and usage in digital financial services. We test two strategies for preventing non-institutional fraud: an anti-fraud campaign and a technical intervention – a unique communications code – which verifies the provenance of messages sent from a digital platform. Using a behavioral laboratory experiment along with survey outcomes, we test whether anti-fraud campaigns increase the ability of MSMEs to distinguish between fraudulent and legitimate communications. Additionally, we test how confidence, trust, and usage of digital financial services respond to anti-fraud campaigns. Finally, we test that the deployment of the unique communications code is suitable for businesses using a follow-up experimental exercise.

Contents

Introduction

Non-institutional fraud (fraud carried out by individuals or groups) targeted at micro, small, and medium enterprises (MSMEs) is pervasive across low- and middle-income countries (LMIC) and has risen in the wake of the COVID-19 pandemic.[1] Not only can fraud lead to immediate (and sometimes severe) monetary and psychological damage, but it can also lead to systemic mistrust in, and underuse of digital services. There is limited knowledge on what mitigation strategies can be taken to reduce fraud and help bring the promise of digital financial services to MSEs in developing countries. This project seeks to address this gap by understanding the impact of i) a learning intervention aimed at consumer capacity to distinguish between fraud and legitimate communication and ii) a unique customer code (UCC) on trust in digital services.

The concept of non-institutional fraud covers a range of potential activities, including phishing[2] 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.[3] At the core, non-institutional fraud is carried out by individuals or groups who are not affiliated with a formal institution (i.e. not insiders in a bank or affiliate) who seek to trick victims into directly sending money, or sending sensitive information that can be used to defraud the victim. Based on the modus operandi of fraudsters, several strategies are often pointed to in preventing its negative ex post effects. We consider two of these strategies: anti-fraud campaigns and a technical intervention – a unique communications code – which verifies the provenance of messages sent from a digital platform. Using this behavioral laboratory experiment, we test whether a simple anti-fraud intervention increase the ability of MSMEs to distinguish between fraudulent and legitimate communications. Likewise, we use this as an opportunity to test the suitability of the deployment of the UCC. Finally, we seek to understand not only the ex-post impacts of these interventions, but also how they might impact ex-ante behavior. Therefore, we study the effect of anti-fraud education on confidence in detecting fraud, trust in, and usage of digital financial services.

Non-institutional fraud is pervasive: IPA's recent consumer protection surveys in Kenya, Nigeria and Uganda found that phishing scams had been faced by 56% of Kenyan respondents, 33% of Ugandan respondents, and 42% of Nigerian respondents. This was the most prevalent issue in Kenya and Uganda, and the third most prevalent in Nigeria.[4] MSEs are common targets of non-

[1] Tade, Oludayo. "COVID-'419': Social Context of Cybercrime in the Age of COVID-19 in Nigeria." *African Security* (2021): 1–24. https://doi.org/10.1080/19392206.2021.2004642.

[2] Phishing refers to 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.

[3] Garz, Seth and Gine, Xavier and Karlan, Dean and Mazer, Rafe and Sanford, Caitlin and Zinman, Jonathan, Consumer Protection for Financial Inclusion in Low- and Middle-Income Countries: Bridging Regulator and Academic Perspectives (December 2020). Global Poverty Research Lab Working Paper No. 20- 110, Annual Review of Economics, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3750236 or http://dx.doi.org/10.2139/ssrn.3750236

[4] Citation pending. From a presentation titled 'Insights from Consumer Surveys in Kenya, Nigeria, & Uganda' delivered in February 2021.

institutional fraud in developing countries, despite perceptions that fraud is targeted at larger businesses.[5] MSEs face fraud risk related to their customers, and also their employees, and have multiple vulnerabilities including business bank accounts, purchases and sales transactions, and business IT infrastructure.

Non-institutional fraud causes immediate and long-term damage. Immediately, fraud leads to monetary loss, but also to psychological impacts including anger, difficulties with trust, feelings of violation, stress, and social embarrassment.[6] In the long-term, low trust may lower willingness to access digital financial services (DFS). This is damaging for MSEs in particular, as digitalization can drive access to market through platform engagement and social media, and access to finance through new digital finance opportunities.[7] Notwithstanding regulatory differences, the lower rate of digital payments in Nigeria compared to Kenya may reflect lower levels of trust in digital financial services.[8]

Conducted in partnership with Amana Market, a digital platform in Nigeria that offers access to market information and financial service to MSEs, this study involves a lab experiment with potential users of the platform. In the lab experiment, MSEs will be randomized into one of four groups: a control, or one of three educational treatment arms. To estimate impact, participants will face the experimental task of successfully distinguishing genuine from fraudulent communication scenarios. We will use baseline and endline survey data to measure a range of outcomes around susceptibility to fraud, trust in platforms, and engagement with the platform. Findings from this study will help improve consumer protection and support digital security for Africa's large and growing market platform[9] and financial services sectors.

We will run the lab experiment in the decision lab that the Busara Center for Behavioral Economics is developing with Ahmadu Bello University (ABU) in Kaduna State, Nigeria. ABU is among the most prominent universities in Nigeria, and we will collaborate with researchers from ABU. Busara and ABU are creating the first decision lab to support behavioral studies run by academics nationally and internationally that allows data collection in an area with high social mixing (by religion, herders vs famers, etc.) and major agricultural production. The lab, which is part of a government research initiative called Nudge Arewa ('North' in Hausa), is affiliated with the Business School as well as the Department of Economics at ABU. The lab infrastructure is located

[5] Salah Kabanda, Maureen Tanner & Cameron Kent (2018) Exploring SME cybersecurity practices in developing countries, Journal of Organizational Computing and Electronic Commerce, 28:3, 269-282, DOI: 10.1080/10919392.2018.1484598

[6] DeLiema, Marguerite and Mottola, Gary R. and Deevy, Martha, Findings from a Pilot Study to Measure Financial Fraud in the United States (February 9, 2017). Available at SSRN: https://ssrn.com/abstract=2914560 or http://dx.doi.org/10.2139/ssrn.2914560

[7] Partnership for Finance in a Digital Africa, "Micro-entrepreneurs in a platform era," Farnham, Surrey, United Kingdom: Caribou Digital Publishing, 2019. https://mse.financedigitalafrica.org.

[8] World Bank Group. 2019. Nigeria Digital Economy Diagnostic Report. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO

[9] Online marketplaces which facilitate commercial transactions between buyers and sellers.

within the Centre of Excellence, a new facility constructed by the Central Bank of Nigeria to encourage collaboration.[10]

Research Questions

There are three core research questions to be answered by the laboratory experiment:

1. Do anti-fraud interventions (like educational trainings or campaigns) increase the ability to distinguish between fraud and legitimate communications, and the confidence in one's ability to do so?
2. Do anti-fraud interventions increase trust and usage of digital financial services?
3. Is the UCC an effective tool in building trust in digital communications? Which approach to the UCC is more effective, pre-specified or personally chosen?
4. Does time pressure increase participants' susceptibility to fraud?

Additionally, we have several supporting research questions including those related to the heterogeneity of effects:

5. Does experience with ICT, DFS, or fraud attempts serve as a substitute or complement to the anti-fraud intervention?
6. Does increased time pressure while identifying fraud attenuate or strengthen the effectiveness of the simple anti-fraud intervention?
7. Is the simple anti-fraud intervention suitably designed for subgroups with different demographic characteristics, such as female-headed businesses, or older business owners?

Research Strategy

Outcome measurement

Ability to distinguish between genuine and fraudulent communications

Participants will be exposed to 20 fictitious scenarios, half fraudulent and half genuine. All participants will be exposed to the same 20 scenarios, with the aim of discerning genuine from fraudulent communications. The scenarios will be shown to each participant in a random order. Our first outcome will be constructed around whether respondents correctly identify these scenarios. These outcomes are listed in table 1.

[10] https://www.abu.edu.ng/news-and-events/news/president-buhari-commissions-cbn-centre-of-excellence-at-abu-zaria.php

Table 1: Outcomes related to the ability to distinguish between genuine and fraudulent communications

Outcome	Tier	Details
Accurate	Primary	A variable indicating if a scenario was correctly identified, regardless of whether that scenario is fraudulent or genuine.
True Positive	Secondary	A variable indicating if a fraudulent scenario was correctly identified
True Negative	Secondary	A variable indicating if a genuine scenario was correctly identified
Confidence	Primary	Self-reported assessment of confidence that they have correctly identified if a scenario is fraudulent or genuine, on a scale of 1 (very unsure)-10 (very confident)

Table 2: Outcomes related to trust and usage of DFS

Outcome	Tier	Details
Likely	Primary	Self-reported likeliness to use DFS in the future, on a scale of 1 (very unlikely)-10 (very likely)
Information	Secondary	Self-reported agreement that their information is kept safe by DFS providers, on a scale of 1 (strongly disagree)-10 (strongly agree)
Money	Secondary	Self-reported assessment of security from fraud when using DFS, on a scale of 1 (very unsecure)-10 (very secure)
Index of trust	Primary	Inverse Correlation Weighted Index of trust outcomes.[11]

Trust and usage of digital financial services

Baseline and endline surveys will be conducted to allow us to test several key research questions, namely, does the financial education intervention in the experiment impact trust in and willingness to use DFS? These outcomes are listed in Table 2.

Recruitment and Sampling

This experiment will be run in partnership with CoAmana, a business in Nigeria which provides services to MSMEs. The sampling frame for the lab experiment will be stratified on the basis of gender, sector (agricultural vs. non-agricultural), and number of staff (1 vs. >1). 750 MSEs will be recruited for this experiment.

Timeline of experimental session

The experiment will begin with a survey to record characteristics of participants, which will take approximately 75 minutes. Next, each participant will be randomized into treatment and comparison group streams, each of which will take part in an information session depending on

[11] Anderson, Michael L. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103, no. 484 (2008): 1481–95. https://doi.org/10.1198/016214508000000841.

the stream allotted to them. These sessions will around 30 minutes for the most intense intervention, with the control group stream taking no time, and the warning group taking about 2 minutes After the information session, participants will complete a fraud identification activity: determining whether presented scenarios are fraudulent or legitimate communications, taking about 15 minutes. Treatment groups from the education component will then additionally be asked about their preferences regarding a hypothetical personal security system – the unique communications code (UCC) component. After this, participants will be brought back to a single stream to take a short survey, which lasts about 15 minutes. Finally, the control group will receive an educational intervention to ensure that all trial participants have the opportunity for the beneficial learning offered by treatment. After three weeks all treatment groups will get a follow up text or phone call with a short survey to measure long term effects of the educational intervention and any effects of the UCC based on their indicated preferences.

Table 3: Anti-Fraud Campaign Interventions

Control	Control group will 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 **+ Simple Warning Message**	On top of the lab manager's session introduction, and the consenting process, T1 subjects will receive on-screen general warning messages stating, "Digital fraud represents a threat to MSEs 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 – over the phone, by email, or in person".
Treatment 2 **+ 7 Key Signs of Fraud**	On top of the lab manager's session introduction, and the consenting process, T2 subjects will receive an on-screen written list of 7 key signs of potential fraud which is narrated in an audio file. This information is prefaced by a general warning message (see Treatment 1). To aid recall, subjects will be 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 will be collected before the remainder of the lab session.
Treatment 3 **+ Illustrative Examples**	On top of the lab manager's session introduction, and the consenting process, T3 subjects will 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 will be prompted to write down the key signs upon completion, before replaying the 7 key signs and filling in the gaps in their answer sheets. Subject's notes will be collected before the remainder of the lab session.

Study Design and Interventions

Interventions

Education Component

All participants are then randomized into either a control group or one of three educational treatment groups. Each of the three treatment groups receives a variation of an educational intervention aiming at helping participants distinguish between genuine and fraudulent communication. These simple educational interventions are meant to replicate common approaches used in anti-fraud campaigns and trainings. 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 3.

Time Component in Fraud Scenarios

Additionally, to test the importance of urgency in correctly identifying scenarios, we will include a time component in half of the scenarios, where participants have less time to respond to the scenario. This is included to mimic a level of stress that people will often face when being exposed to fraud.

Unique Customer Code (UCC) component

After responding to the scenarios, and as part of the endline survey, all participants will be requested to set up a personal security system for them to be able to participate in the final follow up task. We will explain that the purpose of the UCC is to verify the authenticity of future communication coming from us.

Subjects are randomly allocated into one of two equally weighted groups:

- the non-personalized UCC group: these subjects are assigned a randomly generated 5-digit UCC code
- the personalized UCC group: these subjects are instructed to choose their own 5-digit UCC code.

All codes will all be recorded centrally and sent to subjects by SMS to keep as a record. Participants will be instructed to look out for the UCC in future communications from us as proof of authenticity and briefed that they will receive a follow-up SMS in 3 weeks. This follow-up message consists of a simple fictitious scenario, where the sender is Busara, and the participant is the recipient. The scenario will request that recipients confirm their month and year of birth by return text so ensure that it is correctly recorded in Busara's database. The individual UCC is randomly assigned to half of these messages, which will appear otherwise genuine and safe, with no giveaway features other than the presence or absence of the UCC.[12]

[12] A further text will follow upon completion of the exercise thanking participants for their engagement, explaining the purpose of the follow-up exercise, and reminding participants that they should not engage with customer outreach where they are not confident of the authenticity of the sender.

This UCC component will test:

1. How does the presence (absence) of a pre-specified authentication code affect the degree of confidence recipients place in customer outreach (as measured by the response rate)? This allows us to credibly measure whether respondents indeed looked out for the UCC code, which we will use as a proxy for the usefulness of UCC as a security system.

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 (as measured by differential response rates)?

Two weeks after the SMS, a phone call follows with the purpose of:

- Encouraging non-responders to respond to the SMS scenario
- Conducting a quiz centered upon recall of the key signs of potential fraud (3 multiple choice questions)[13] This will allow us to test the degree of decay in knowledge retention and observe whether this varies in accordance with the intensity of the original educational intervention.
- Posing a final question about subjects' preferences regarding future UCC formatting (i.e., preference over (a) numerical code, (b) word, or (c) sentence or phrase.

Study Hypotheses

This study will explore whether providing MSEs with education about digital fraud affects their ability to distinguish fraudulent and genuine communications, their confidence in doing so, their trust in digital financial services, and their willingness to use digital financial services. These are reflected in detail in hypotheses 1.0-1.4, 2.0, 3.0, and 4.0 in table 4. Additionally, we will explore heterogeneity by treatment for confidence, trust, and usage outcomes. The study will additionally gauge preferences regarding the format of a hypothetical personal security system (i.e., unique code, unique phrase, unique image) that would be embedded in future communications from digital financial service providers as a digital communication authentication tool. These are reflected in hypotheses 5.1-5.3 in table 4.

Balance Checks

We will test that those who are assigned to treatment are not different from those who are assigned to control. We plan to use a joint test of orthogonality to test balance across treatment groups, holding out those variables that we have already stratified treatment upon (e.g., gender and occupation). Additionally, since there are multiple treatment groups within our experiment, we will perform a multinomial logit regression and then test for joint orthogonality of coefficients.

[13] A similar short quiz consisting of 3 multiple choice questions will be administered at the close of the endline survey to give us two data points with which to evaluate decay.

Table 4: Research hypotheses for core research questions

Research Question	Number	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?

How does urgency impact confidence and ability to detect fraud?	6.1	When asked to identify a fraudulent scenario under time pressure, does this reduce confidence and ability to do so?
	6.2	Are anti-fraud interventions attenuated by time pressure, or do they protect against this kind of time pressure?
Is knowledge from educational interventions effectively retained over a short time horizon?	7.1	Is there evidence of decaying performance between knowledge retention quizzes administered immediately after intervention, and at +3 weeks?
	7.2	Does the rate of knowledge decay vary in accordance with the intensity of the original educational intervention administered?

Statistical Power

To assess the sample size requirement for the lab experiment, we estimate the minimum detectable effects (MDE) under 32 alternative design scenarios (varying by sample size, power, and number of treatment arms) in Table 5.

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).

Table 5: Estimated statistical power under alternative design scenarios (power, number of treatment arms, sample size)

Outcome: Detection accuracy									
		Power 90%				**Power 80%**			
MDE	**N**	**1 Tr.**	**2 Tr.**	**3 Tr.**	**4 Tr.**	**1 Tr.**	**2 Tr.**	**3 Tr.**	**4 Tr.**
	250	20.58%	25.23%	29.22%	32.73%	17.79%	21.81%	25.25%	28.29%
	500	14.52%	17.82%	20.58%	23.02%	12.55%	15.40%	17.79%	19.91%
	750	11.85%	14.52%	16.78%	18.78%	10.24%	12.55%	14.50%	16.23%
	1000	10.26%	12.57%	14.52%	16.25%	8.87%	10.86%	12.55%	14.04%
Note: power calculations assume a baseline detection ability of 50% (chance), with a corresponding standard deviation of 50%.Tr. = Treatment.									

To give an 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.

In our chosen experimental design, we will include 750 participants, and 3 treatment arms, giving us a minimum detectable effect size of 14.50% (at 80% power) or 16.78% (at 90% power). This gives an average of 187.5 participants per cell - including a control and 3 treatment arms.

Empirical Strategy

Estimation of treatment effects

Fraud Scenarios

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

$$Y_{is} = \alpha + \beta_1 Warning_i + \beta_2 T_i + \beta_3 T_i \times Vingettes_i + \varepsilon_i$$

where we define Y_{is} to be one of the outcome variables described in table 1 or table 2 for participant i (and scenario s), $Warning_i$ is an indicator for the general warning message regarding fraud, T_i is an indicator for the seven key signs, $Vingettes_i$ an indicator for if applied illustrative examples are used. β_1 estimates the treatment effect of treatment 1 (general warning message), β_2 estimates the treatment effect of treatment 2 (seven key signs by text/audio), β_3 estimates the treatment effect of treatment 3 (addition of illustrated examples by text/audio).

To test the hypotheses above, we will perform the following hypothesis tests after running the regression using accuracy (i.e., accurate identification of fraud scenarios) as an outcome.

- **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:** $H_0: \beta_2 \leq 0$. Providing MSEs with seven key warning signs for potential fraud in a simple format (written/audio) 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).

- **Hypothesis 1.3:** $H_0: \beta_3 \leq 0$. Illustrating applied examples of fraudulent communications in a simple format (written/audio) improves MSEs' ability to distinguish between genuine and fraudulent communications, still further than can be achieved with simple warning signs alone (T3 vs. T2).

The other hypotheses relating to confidence, trust of DFS, and likelihood of using DFS will be tested in the same manner. We will also test if any of the treatments will improve the ability to distinguish between genuine and fraudulent communications (i.e., hypothesis 1.0 in table 4) by running a restricted specification:

$$Y_{is} = \alpha + \beta_1 T_i + \varepsilon_i$$

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

Trust, Usage, and Confidence

Given the degree of precision in their elicitation, our main specifications for outcomes related to trust, usage, and confidence will be estimated as the restricted specification for the fraud scenarios. For the outcome of confidence in distinguishing fraudulent scenarios, we estimate:

$$Y_{is} = \alpha + \beta_1 T_i + \varepsilon_i$$

where Y_{is} is the confidence is participant i's assessment of their confidence in correctly identifying scenario s as fraudulent or legitimate? For respondents' assessment of the likelihood of trust and usage in digital financial services in the future, we estimate a similar specification, except s now indexes the type of digital financial services. For these outcomes, we will also explore heterogeneity by treatment.

However, the outcomes for trust, usage, and confidence are not cardinal but ordinal. Therefore, as a robustness test for these outcomes, we will estimate an ordered logit or ordered probit regression to test for treatment effects. Finally, while the ordered dependent outcomes will be used for our main hypothesis testing, we may transform these outcomes to binary (for example, by taking outcomes above median) in order to generate more easy-to-understand estimates to communicate with policymakers.

Time Pressure

Some of the fraud scenarios will include a time pressure component, to replicate the kind of urgency that scams are often delivered with. This will be random by scenario. We can estimate the effect of urgency on accuracy using the following specification:

$$Y_{is} = \alpha + \gamma_1 Urgency_i + \varepsilon_i$$

where $Urgency_i$ indicates a timer appeared during the experimental task. We test the null $H_0: \gamma \geq 0$. Additionally, given that urgency impacts the success of the treatment we will employ a specification that more finely captures the level of urgency:

$$Y_{is} = \alpha + \beta T_i + \gamma_1 \times Urgency_{1i} + \gamma_1 \times Urgency_{2i} + \gamma_1 \times Urgency_{3i} + \varepsilon_i$$

where the three levels of urgency feature a timer of 30 seconds, 45 seconds, and 60 seconds to complete the experimental task, respectively.

Heterogeneous Treatment Effects

Experience with ICT, DFS, and Fraud

Does experience with DFS and fraud serve as a substitute or complement to the anti-fraud campaign? The baseline survey will collect information relating to the participants' level of experience with information communication technology (ICT) and DFS, as well as exposure to fraud and scams in ICT and DFS, and levels of trust.

For many of these outcomes, a standardized index will be computed then split into types by those who are above or below average according to that index. Indices will tend to be computed using Principal Components Analysis and taking the first component of that index. Where noted below we may use a context specific index (see, for example, Fraud Experience).

Time Pressure

We are also interested in how urgency interacts with the anti-fraud intervention. Our main specification to test this hypothesis is:

$$Y_{is} = \alpha + \beta T_i + \gamma Urgency_i + \delta T_i \times Urgency_i + \varepsilon_i$$

This specification will be preferred in cases where either power is limited for analysis of heterogeneity or there is limited theory of change to support the fully interacted specification. We test $H_0: \delta = 0$. In addition, given sufficient power, we may extend this test to levels of urgency or treatment arm, though likely not both.

Demographic Characteristics

To allow for segmentation and analysis of heterogeneity, the survey will additionally collect information relating to attitudinal and behavioral characteristics, as well as relevant demographic factors.

Table 6: Heterogeneous Treatment Effects

Variable of interest	Details
ICT experience	Standardized index of experiences with information communication technologies. After indexing, individuals will be split into high and low experience types.
DFS experience	Standardized index of experiences with digital financial services. After indexing, individuals will be split into high and low experience types.
Fraud experience	Respondents will be split into as many as four types, conditional on the underlying data: 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. For sake of power, we may reduce these categories to as few as two. For example, if few people responded or faced losses due to fraud, we will reduce to those who have and have not encountered fraud.
Gender	An indicator variable equal to one if the business owner is a woman, zero otherwise.
Age	An indicator variable for if the business owner is above (or below) the median age.
Occupation	A set of indicator variables (and a left-out group) for the following occupations: ● Agriculture ● Non-Agriculture ● Student
Self-Control	A standardized index of self-control, impulsiveness, attentiveness. After indexing, individuals will be 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.[14] After indexing this may be split into high and low risk types.
Generalized Trust and Skepticism	A standardized index of variables associated with generalized social trust and skepticism, including questioning mind (Fullerton and Durtschi, 2004).[15] After indexing, participants will be split into a high and low trust types.

[14] A considerable number of participants will be Muslim. There is a chance these questions do not provide substantial variation for these participants if they are associated by these respondents with gambling, which is *haram*, or forbidden, in Islam. We have written the questions to protect against this outcome, framing them in a business context, but if we are not confident that we succeeded, we may elect to omit these results.

[15] Fullerton, R., & Durtschi, C. (2004). The Effect of Professional Skepticism on the Fraud Detection Skills of Internal Auditors. https://doi.org/10.2139/ssrn.617062

UCC follow-up exercise

The analysis of the choice experiment will look to see the overall feasibility of the UCC. First, we will check how often people responded to the messages with and without their UCC included. We estimate this specification:

$$Y_i = \alpha + \eta\, UCC_i + \varepsilon_i$$

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 $H_0: \eta \leq 0$. Additionally, within the choice experiment we will explore heterogeneity by whether the UCC was automatically assigned or chosen by the participant themselves:

$$Y_i = \alpha_1 + \alpha_2\, UCC_i + \ \alpha_3\, UCC_i \times Personalized_i \ + \varepsilon_i$$

We test H_0: $\alpha_3 \leq 0$.

Learning by Doing

In addition to considering heterogeneity in learning effects by fraud experience, we will explore learning by doing within the experiment. A simplest specification for testing for learning by doing is to simply test if those scenarios that appeared later in the order were more often correctly identified by participants. We specify this,

$$Y_{is} = \alpha + \theta Order_{is} + \varepsilon_i$$

where $Order_{is}$ is the order scenario s was presented to participant i. We test $H_0: \theta \leq 0$ to test for learning. Alternatively, given that learning may be non-linear, we can approach estimation using a fully saturated model:

$$Y_{is} = \alpha + \theta_1 First_{is} + \cdots + \theta_{20} Twentieth_{is} + \varepsilon_i$$

where $First_{is}$ is a variable indicating if scenario s was first in the order for participant i, for example. For this estimation we jointly test $H_0: \theta_n \leq 0$ for $n = 1, \ldots, 20$.

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.[16]

[16] For robustness, we could also cluster at the experimental session level.

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).[17] Additionally, where appropriate and for purposes of robustness, we will include family wise error rate (FWER)-adjusted p-values and False Discovery Rate (FDR)-adjusted p-values.

Fieldwork

Data collection

Prior to the experiment, we organized a 3-day pilot in which we made sure that modules and the experimental protocol ran smoothly and adjusted based on any issues faced. We made note of all such adjustments. The data from the pilot will not be used in the main analysis.

We expect the data collection process to take 7 weeks excluding the pilot. Lab managers reviewed and were trained on the survey before the pilot and received a refresher training one day before the first pilot and one additional update after the pilot and for smooth execution of the modules and experiment. There will be a total of 2 lab sessions daily. There will be 15 participants per lab session, making it possible to survey 30 participants in a day. Hence surveying 750 participants will take 25 days and we keep 10 days as a buffer in case of unforeseen events.

Data from the experiment will be sent directly from participants' tablets to the research teams' computers, and then downloaded via .csv output from the Survey CTO platform, through which the modules are run. The data will be kept anonymous and hosted on a folder shared among the research team. Data protection procedures have been approved by both Ahmadu Bello University's IRB and Trinity College Dublin's IRB.

Data management

Survey response data collected using the SurveyCTO platform will first be stored on android phones used by enumerators. All data will be encrypted on the tablets upon completion of the survey. Data will automatically be compiled onto the Busara server, where the research team can download it. The data will then be cleaned using a statistical analysis program. All respondents will be given a unique identification code, and all Personally Identifiable Information (PII) will be stripped from the main dataset and saved in separate files from the survey responses. Once compiled and de-identified, both the raw and clean data will be backed up and stored on external hard drives managed by the research team. All compiled raw and clean data stored on senior staff computers, or the back-up external drive will be encrypted prior to storage using the

[17] *Ibid.*

TrueCrypt software. Further, all existing and new Busara staff have been extensively trained in the importance of keeping all data confidential.

The researchers on the research team for the proposed project who analyze the data collected will do so using a dataset that is de-identified in the sense that the names and any other personal identifiers (such as phone numbers) of respondents are removed. No information that can be traced back to an individual will be presented in reports, articles or otherwise public documents produced during this study. Participation lists will be stored in a locked office in a locked file cabinet or on a password-protected computer and will be used for administrative purposes.

Personal data - including interview recordings in SurveyCTO - will be held by Busara for the duration of the project (12 months). During this period, it will be encrypted using TrueCrypt and only used to verify incentives or other key information. The personal data will not be shared with Co-Investigators outside of Busara. Personal data will not be saved locally. Analysis on the data will only be conducted on the anonymized data and will work on computers that are password protected. The personal data will be deleted from its encrypted form to destroy it.

In accordance with the open access policy of the Gates foundation, the data will be made public after stripping any personal identifying information.[18] The study is funded by a $350,000 USD grant from the Innovation for Poverty Action Initiative on Consumer Protection, a program funded by the Gates Foundation.

[18] https://www.gatesfoundation.org/about/policies-and-resources/open-access-policy