

Can you spot a scam?

Measuring and improving scam identification ability

Elif Kubilay¹ (r) Eva Raiber² (r) Lisa Spantig^{3,1} (r) Jana Cahlíková⁴ (r) Lucy Kaaria⁵

¹Essex ²Aix-Marseille ³Aachen ⁴Bonn ⁵Nairobi

EEA

August 2023

Motivation

- Expansion of digital financial services in developing countries
 - increases access to finance (e.g., Pazarbasioglu et al. 2020, Balyuk 2022)
 - increases consumer protection issues (Garz et al. 2021): high and hidden prices, over-indebtedness, post-contract exploitation, fraud, ...

Motivation

- Expansion of digital financial services in developing countries
 - increases access to finance (e.g., Pazarbasioglu et al. 2020, Balyuk 2022)
 - increases consumer protection issues (Garz et al. 2021): high and hidden prices, over-indebtedness, post-contract exploitation, fraud, ...
- We focus on fraud. Costly in several ways
 1. direct monetary costs for victims

Motivation

- Expansion of digital financial services in developing countries
 - increases access to finance (e.g., Pazarbasioglu et al. 2020, Balyuk 2022)
 - increases consumer protection issues (Garz et al. 2021): high and hidden prices, over-indebtedness, post-contract exploitation, fraud, ...
 - We focus on fraud. Costly in several ways
 1. direct monetary costs for victims
 2. indirect costs
 - loss of confidence in financial matters (Brenner et al. 2020)
 - erosion of trust in financial institutions (Johnson et al. 2019)
- individuals might not take up or use digital financial services

Motivation

- Expansion of digital financial services in developing countries
 - increases access to finance (e.g., Pazarbasioglu et al. 2020, Balyuk 2022)
 - increases consumer protection issues (Garz et al. 2021): high and hidden prices, over-indebtedness, post-contract exploitation, fraud, ...
- We focus on fraud. Costly in several ways
 1. direct monetary costs for victims
 2. indirect costs
 - loss of confidence in financial matters (Brenner et al. 2020)
 - erosion of trust in financial institutions (Johnson et al. 2019)
 - individuals might not take up or use digital financial services
- We focus on a large digital financial market: Kenya (Koyama et al 2021)

Fraud in Kenya

- Predominant type of fraud in Kenya: phone scams (Blackmon et al. 2021)
- 56% reported they had been contacted by scammers in the past six months (Blackmon et al. 2021)
- 90% of the adult population is concerned about fraud when using digital services (Koyama et al. 2021)
- 71% of the self-employed report limiting their usage of DFS due to concerns about fraud (Koyama et al. 2021)

Prevention of Victimization

- A common approach to tackle fraud:
Education and awareness campaigns

3

Never Share Your PIN/Password
Nobody should know your PIN or password, including merchants, relatives, and friends. Sharing the information can cause unauthorised access to your account.

#SecureYourBanking KAA CHONJO! Learn

SECURE YOUR BANKING
MOBILE, CARD & ONLINE SAFETY

Contact your bank for more information on card and online safety. FOLLOW US ON TWITTER @KenyaBankers #KaaChorjo

KENYA BANKERS ASSOCIATION

Paystack Link
Easy. Affordable. Instant.

Credit Bank
The Power of Your Bank

Prevention of Victimization

- A common approach to tackle fraud: Education and awareness campaigns
- How effective are they?



3

Never Share Your PIN/Password
Nobody should know your PIN or password, including merchants, relatives, and friends. Sharing the information can cause unauthorised access to your account.

#SecureYourBanking KAA CHONJO! Learn

SECURE YOUR BANKING
MOBILE, CARD & ONLINE SAFETY

Contact your bank for more information on card and online safety. FOLLOW US ON TWITTER @KenyaBankers #KaaChorjo

KENYA BANKERS ASSOCIATION

Paystack Link
Easy. Affordable. Instant.

Credit Bank
The Future of Banking

This Study

Research Questions

- How well can individuals distinguish genuine from fraudulent contact attempts?
Who is more susceptible to scams?
- How confident are individuals in their scam identification ability?
Who is very confident?
- Can a light-touch educational intervention increase scam identification ability?

This Study

Research Questions

- How well can individuals distinguish genuine from fraudulent contact attempts?
Who is more susceptible to scams?
- How confident are individuals in their scam identification ability?
Who is very confident?
- Can a light-touch educational intervention increase scam identification ability?

Steps

1. Scoping: stakeholder interviews, focus groups, social media analysis
2. Develop a measure of scam identification ability and implement it in an online survey
3. Test if information provision improves this ability

This Study

Research Questions

- How well can individuals distinguish genuine from fraudulent contact attempts?
Who is more susceptible to scams?
- How confident are individuals in their scam identification ability?
Who is very confident?
- Can a light-touch educational intervention increase scam identification ability?

Steps

1. Scoping: stakeholder interviews, focus groups, social media analysis
2. Develop a measure of scam identification ability and implement it in an online survey
3. Test if information provision improves this ability

Main Take-Aways

- women and less experienced users are more susceptible
- a light-touch educational intervention does not increase scam identification ability, but it makes individuals more cautious

Contributions

1. Financial fraud in developing countries

Annan (2022a, 2022b), Garz et al. (2021), Blackmon et al. (2021), Fu and Mishra (2022)

→ Measure of *relative* ability to distinguish scams and genuine official communication

2. Approaches to fraud prevention

e.g., Burke et al. (2022), Scheibe et al. (2014), Sheng et al. (2007)

→ Test common policy response in a digital financial market

3. Correlates of fraud susceptibility and victimization

e.g., Moustafa et al. (2021), Norris et al. (2019), Engels et al. (2020), Chen et al. (2018)

→ Document that more susceptible groups do not differentially benefit from education

Data collection

Scoping

Overview of Activities

- Qualitative data: understand perceptions of scams
 - 6 interviews with stakeholders
 - 5 focus group discussions with DFS users
- Social media data: examples of scams and genuine messages
 - Twitter: 427,121 original post; public facebook groups: 18,022 posts
 - survey in Kenya's largest fraud-detection facebook group: 919 responses [more](#)
 - analysis: topic clustering and manual classification of 1,836 examples regarding [topics](#) , [actions](#) , [senders](#)

Main insights

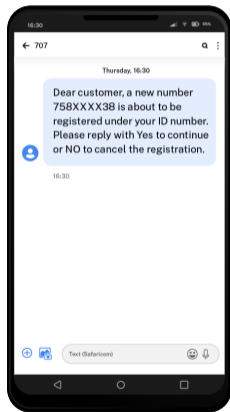
- Scam Types
 - Impersonation of agents/family/friends
 - False loan or investment offers/promotions/prizes
 - Erroneous transfer/shipment
- Goals of scammers: acquire personal info, money, or identity

Measure of Scam Identification Ability

- labeled Twitter data and survey data of examples for scams and official messages ($N = 1,836$)
 - keep only copy-pasted text or text extracted from pictures
- Idea: generate variation by choosing rather hard-to-classify messages
 1. build database of more ambiguous messages
 2. stratified by topic, randomly select 13 scam and 7 official messages examples
 3. pilot in 2 convenience samples ($N = 39$)
 4. final selection based on topics, classification, and certainty rating:
8 scam and 4 official messages examples
 5. vary the display of the sender

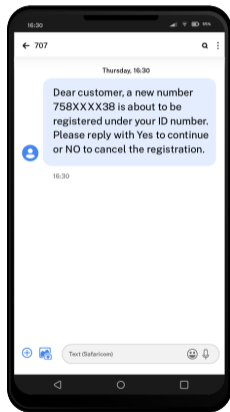
Measuring Scam Identification Ability & Confidence

- Is this a scam message? (Yes/No)



Measuring Scam Identification Ability & Confidence

- Is this a scam message? (Yes/No)



- How confident are you in your answer? (Scale from 1 to 5)

Overview of Vignettes

| | Content | Intention | Sender |
|---------|---|------------|---------------|
| Block A | M-PESA transfer receipt | Genuine | Displayed |
| | Offer to use the new M-PESA app and get cash back | Genuine | Not displayed |
| | Random message to encourage contact | Fraudulent | Displayed |
| | Investment opportunity | Fraudulent | Displayed |
| | Suspended bank account | Fraudulent | Not displayed |
| | Notification as emergency contact | Fraudulent | Not displayed |
| Block B | M-PESA reversal request | Genuine | Displayed |
| | Notification of new registered SIM | Genuine | Displayed |
| | Job offer | Fraudulent | Displayed |
| | Lottery win | Fraudulent | Displayed |
| | Covid-19 relief fund | Fraudulent | Not displayed |
| | Notification as loan grantor | Fraudulent | Displayed |

Implementation

- Online experiment
 - 1000 Kenyan respondents from a consumer panel of Geopoll
 - Quotas for gender, age, and location
 - Survey questions: demographics, the use of DFS, and scam experiences

Implementation

- Online experiment
 - 1000 Kenyan respondents from a consumer panel of Geopoll
 - Quotas for gender, age, and location
 - Survey questions: demographics, the use of DFS, and scam experiences
- SIA measurement
 - 2 blocks with 4 scam messages and 2 official messages each
 - random order of blocks and messages within blocks
 - share of correctly classified messages
 - share of correctly classified scams
 - share of correctly classified non-scams

Implementation

- Online experiment
 - 1000 Kenyan respondents from a consumer panel of Geopoll
 - Quotas for gender, age, and location
 - Survey questions: demographics, the use of DFS, and scam experiences
- SIA measurement
 - 2 blocks with 4 scam messages and 2 official messages each
 - random order of blocks and messages within blocks
 - share of correctly classified messages
 - share of correctly classified scams
 - share of correctly classified non-scams
- Logistics
 - Participation fee: 500 KES (4.4 USD)
 - Average duration: about 21 minutes
 - AEA Registry No: AEARCTR-0008754
 - January 2022

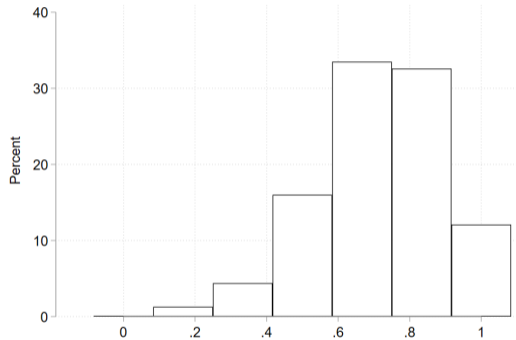
Descriptive Statistics

Descriptive Statistics

| | N | Mean | SD | Min. | Max. |
|---|------|-------|------|------|------|
| Demographics | | | | | |
| Female (0/1) | 1000 | 0.50 | 0.50 | 0 | 1 |
| Age | 1000 | 32.28 | 9.84 | 18 | 67 |
| Urban (0/1) | 999 | 0.50 | 0.50 | 0 | 1 |
| Post secondary education (0/1) | 1000 | 0.73 | 0.44 | 0 | 1 |
| Low income | 1000 | 0.78 | 0.41 | 0 | 1 |
| Formal employment (0/1) | 997 | 0.36 | 0.48 | 0 | 1 |
| Internet on phone (0/1) | 999 | 0.99 | 0.09 | 0 | 1 |
| Social media on phone (0/1) | 999 | 0.99 | 0.09 | 0 | 1 |
| Financial transactions w/ phone in the past 90 days | 980 | 0.96 | 0.21 | 0 | 1 |
| DFS Use | | | | | |
| Number of DFS used | 1000 | 4.78 | 2.52 | 0 | 9 |
| Scam Experience | | | | | |
| Have you ever been contacted by a scammer? | 999 | 0.96 | 0.18 | 0 | 1 |
| Ever been a victim of a scammer? | 960 | 0.56 | 0.50 | 0 | 1 |
| Anyone you know ever been a victim of a scammer? | 1000 | 0.85 | 0.35 | 0 | 1 |
| Scam Identification Ability (Block 1) | | | | | |
| Share of correctly identified messages (SIA) | 1000 | 0.71 | 0.18 | 0 | 1 |
| Share of correctly identified scams | 1000 | 0.74 | 0.24 | 0 | 1 |
| Share of correctly identified non-scams | 1000 | 0.66 | 0.35 | 0 | 1 |
| Average confidence in SIA | 1000 | 4.23 | 0.63 | 1 | 5 |

Heterogeneity in SIA and Confidence

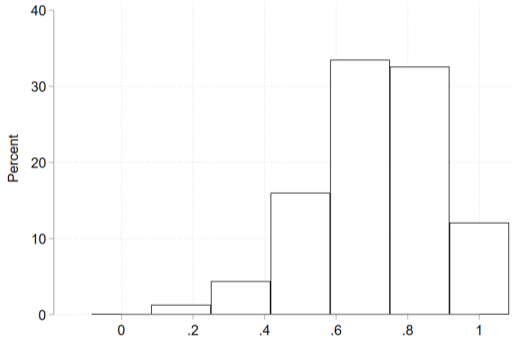
Block 1



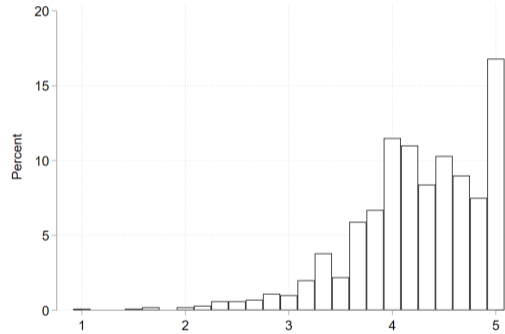
mean SIA=0.71 or 4.23 messages

Heterogeneity in SIA and Confidence

Block 1



mean SIA=0.71 or 4.23 messages



mean confidence=4.23

Who is doing better? Who is more confident?

Block 1

Who is doing better? Who is more confident?

Block 1

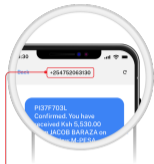
| | SIA | | | Confidence in SIA | | |
|------------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| Demographics: | | | | | | |
| Female | -0.03*** (0.01) | -0.03*** (0.01) | -0.03*** (0.01) | -0.11*** (0.04) | -0.10*** (0.04) | -0.12*** (0.04) |
| Age in Years | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) | 0.01** (0.00) | 0.01** (0.00) | 0.00* (0.00) |
| Post-Secondary Education | 0.03* (0.01) | 0.02 (0.01) | 0.01 (0.01) | 0.11** (0.05) | 0.10* (0.05) | 0.13** (0.06) |
| Low Income | 0.02 (0.01) | 0.02 (0.01) | 0.02 (0.01) | 0.04 (0.05) | 0.03 (0.05) | 0.04 (0.05) |
| Formal Employment | -0.00 (0.01) | -0.01 (0.01) | -0.00 (0.01) | 0.06 (0.04) | 0.04 (0.04) | 0.03 (0.04) |
| DFS Use: | | | | | | |
| Low Trust in DFS | | 0.01 (0.01) | 0.01 (0.01) | | -0.11** (0.05) | -0.09** (0.04) |
| Above average use of different DFS | | 0.03** (0.01) | 0.03** (0.01) | | 0.05 (0.04) | 0.03 (0.04) |
| Scam Experience: | | | | | | |
| Contacted less than 1 week ago | | | -0.01 (0.02) | | | -0.01 (0.06) |
| Victim of a Scammer | | | -0.01 (0.01) | | | -0.05 (0.04) |
| N | 997 | 997 | 956 | 997 | 997 | 956 |
| R-Squared | 0.05 | 0.05 | 0.05 | 0.03 | 0.04 | 0.04 |

Effects of Scam Education

Light-Touch Scam Education

- After completing block 1, 50% of participants receive a light-touch scam education
- Consumers are encouraged to look for **scam markers**
 - typos and grammar mistakes
 - an unknown sender
 - a shortened link
 - requests for private information such as PIN codes or passwords
- Based on Kenyan information campaigns
- Randomized at the individual level
- On average, participants spend more than one minute reviewing the information

Scam Identification Tips



Pay attention to the sender!

- Do you recognize the sender?



Pay attention to the sender!

- Do you recognize the sender?
- Safaricom will only SMS you from M-PESA and Safaricom.



Pay attention to the text!

- Beware of spelling mistakes, wrong tense or wrong punctuation.



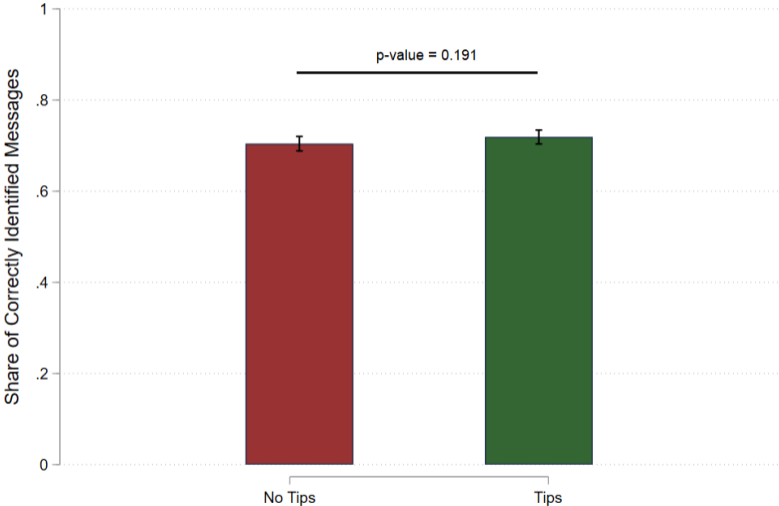
Pay attention to the text!

- Beware of spelling mistakes, wrong tense or wrong punctuation.
- Do not click on shortened links.



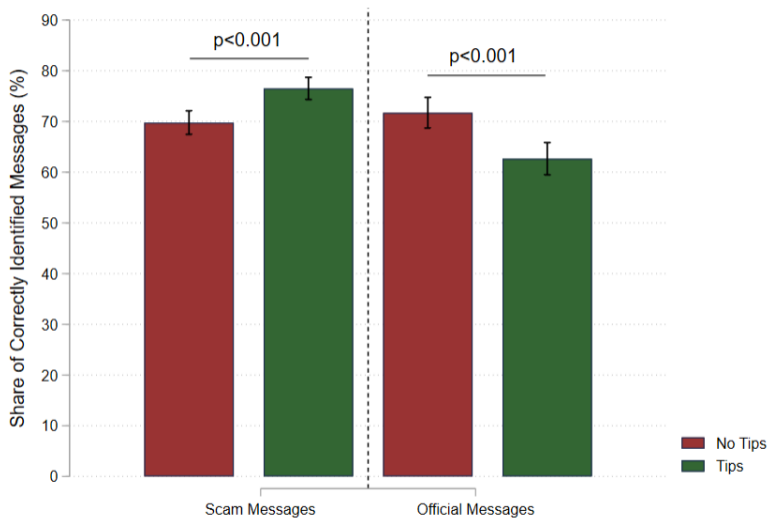
Your bank will never text to ask for your PIN or password!

Tips don't increase scam identification ability



Share of correctly identified messages in block 2

Tips make participants more likely to say a given message is scam



Additional results

- Little treatment effect heterogeneity in SIA
 - individuals with higher education benefit from tips [more](#)
- Tips make participants more confident
 - driven by an increase in confidence for scam messages
- No treatment effect heterogeneity in confidence [more](#)

Additional results

- Little treatment effect heterogeneity in SIA
 - individuals with higher education benefit from tips [more](#)
- Tips make participants more confident
 - driven by an increase in confidence for scam messages
- No treatment effect heterogeneity in confidence [more](#)
- Robustness
 - slightly larger effects when excluding those who failed attention check [tables](#)
 - control variables are balanced across treatments [tables](#)
 - no effect of inclusion/exclusion of control variables [table](#)

Discussion

Discussion

1. Do individuals provide effort in our measure?
2. Why do individuals become more likely to say any given message is a scam?
3. How to think about the effect sizes?

Do individuals provide effort? Incentive treatment

- Additional payment (10 KSH) for each correctly classified message
- For all 12 vignettes
- Randomized at the individual level
- Feedback at the very end of the survey

Do individuals provide effort? Incentive treatment

- Additional payment (10 KSH) for each correctly classified message
- For all 12 vignettes
- Randomized at the individual level
- Feedback at the very end of the survey

Effect of Incentives in Block 1

| | SIA | Scams Identified | Non-scams Identified | Confidence |
|--------------|-----------------|------------------|----------------------|----------------|
| Incentives | -0.01 (0.02) | -0.02 (0.02) | 0.00 (0.03) | 0.06 (0.06) |
| Control Mean | 0.71 | 0.75 | 0.63 | 4.23 |
| N | 956 | 956 | 956 | 956 |
| R-Squared | 0.05 | 0.04 | 0.09 | 0.04 |

Notes: Coefficients from OLS regressions and robust standard errors in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

No difference between tips with incentives and without

Treatment Effects in Block 2

| | Correctly Identified Messages | | | Confidence | | |
|-------------------------------|-------------------------------|-------------------|--------------------|-------------------|-------------------|----------------|
| | SIA | Scams | Non-scams | SIA | Scams | Non-scams |
| Tips (unincentivized) | 0.02 (0.02) | 0.08*** (0.02) | -0.09*** (0.03) | 0.13*** (0.04) | 0.16*** (0.05) | 0.07 (0.06) |
| Tips (incentivized) | 0.03* (0.02) | 0.08*** (0.02) | -0.07** (0.03) | 0.09** (0.04) | 0.09* (0.05) | 0.09 (0.06) |
| Control Mean | 0.70 | 0.69 | 0.71 | 4.20 | 4.20 | 4.33 |
| p-value ($Tips^U = Tips^I$) | 0.56 | 0.89 | 0.40 | 0.43 | 0.18 | 0.78 |
| N | 956 | 956 | 956 | 956 | 956 | 956 |
| R-Squared | 0.04 | 0.10 | 0.16 | 0.46 | 0.40 | 0.26 |

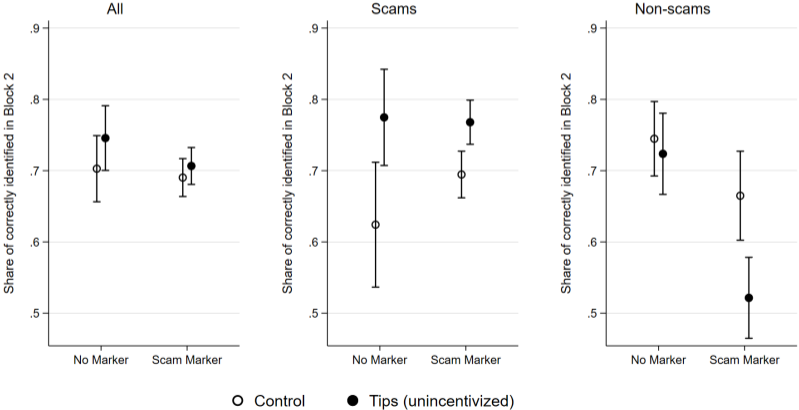
Notes: The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Why do individuals become more likely to say any given message is a scam?

Vignette-level effects by whether the message contains a scam marker

Why do individuals become more likely to say any given message is a scam?

Vignette-level effects by whether the message contains a scam marker



Interpretation of effect sizes

- Absolute levels of SIA are hard to interpret
 - would need to capture all messages and their frequency
 - abstract away from situational circumstances
- focus on differences between groups
 - treatment groups
 - socio-demographics

Interpretation of effect sizes

- Absolute levels of SIA are hard to interpret
 - would need to capture all messages and their frequency
 - abstract away from situational circumstances
 - focus on differences between groups
 - treatment groups
 - socio-demographics
- Magnitude of treatment effects
 - upper bound: literate and relatively educated sample, tips are provided when needed and in a salient way
 - lower bound: participants are alert, use of common tips, difficulty of vignettes [more](#)
 - focus on the direction of the treatment effects

Conclusion

Conclusion

Education Campaigns

- Popular policy response
- No significant effects of information on correctly identified messages
- Respondents become overly cautious: the share of messages categorized as scams increases

Conclusion

Education Campaigns

- Popular policy response
- No significant effects of information on correctly identified messages
- Respondents become overly cautious: the share of messages categorized as scams increases

Several reasons why information might not be effective

- Official communication includes scam-like features
- Not targeted at specific audiences
- Static

Conclusion

Education Campaigns

- Popular policy response
- No significant effects of information on correctly identified messages
- Respondents become overly cautious: the share of messages categorized as scams increases

Several reasons why information might not be effective

- Official communication includes scam-like features
- Not targeted at specific audiences
- Static

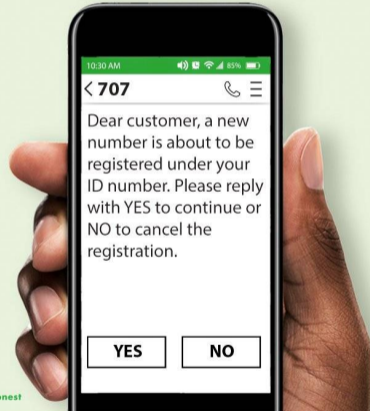
Can we find a better way to increase DFS use sustainably?

Thank you!

lisa.spantig@essex.ac.uk

LET'S PUT A STOP TO FRAUDULENT SIM REGISTRATION

#TUWAANIKE

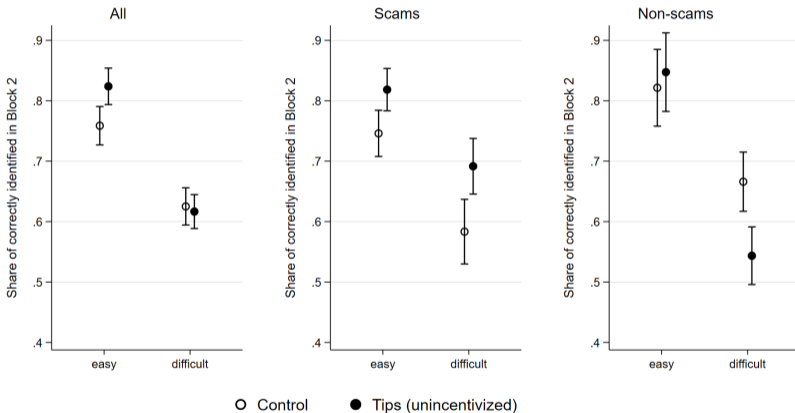


Simple • Transparent • Honest

Appendix

Exploring the difficulty of the vignette

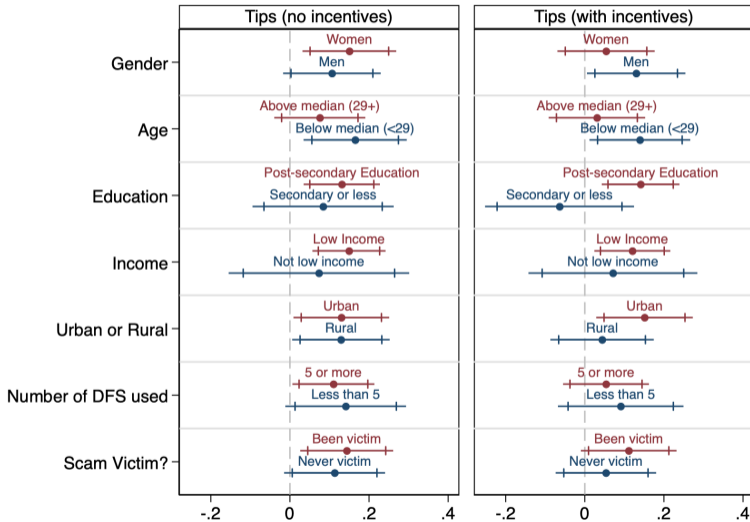
Vignette-level effects by baseline difficulty of the vignette



Treatment Effect Heterogeneity: SIA [back](#)



Treatment Effect Heterogeneity: Confidence [back](#)

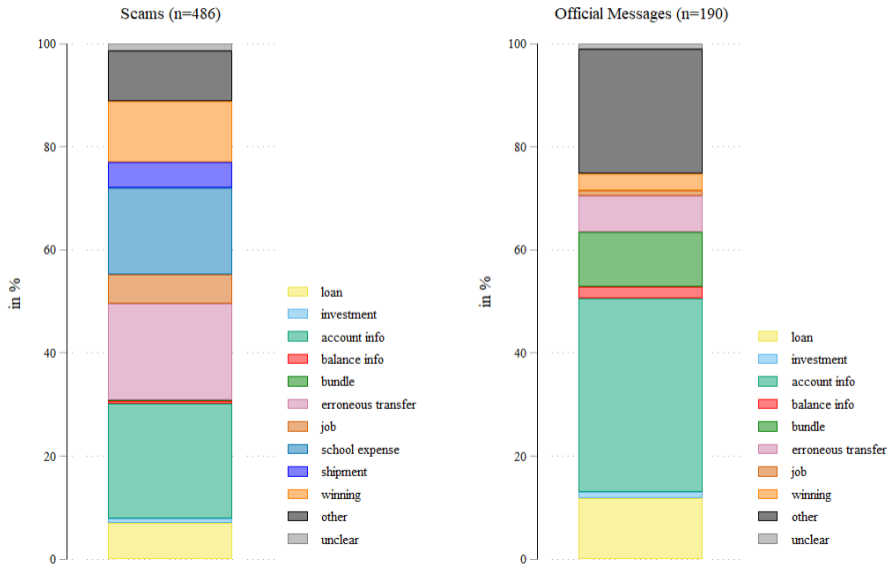


Scoping: Social Media Analysis

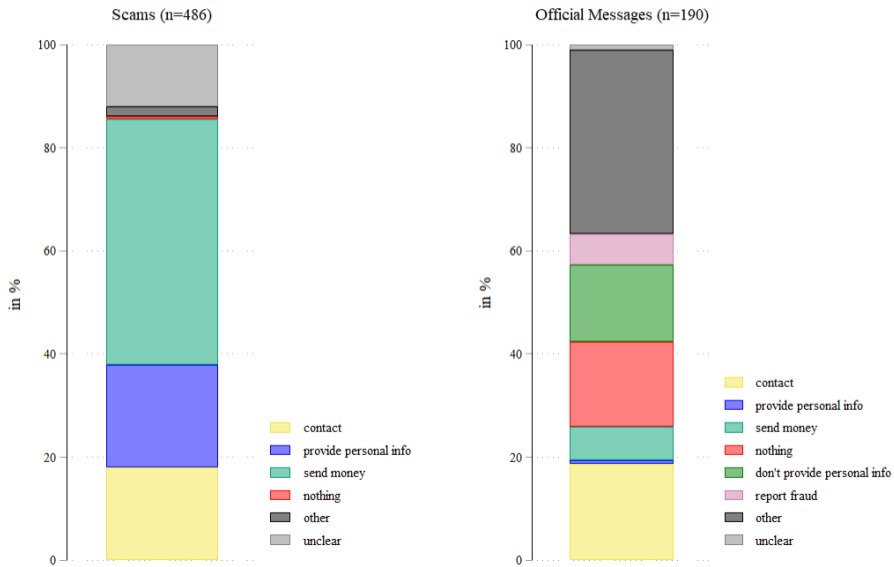
- Twitter (via brandwatch)
 - January 2020 - August 2021
 - 427,121 posts
- Public facebook groups (via crowdtangle)
 - June 2020 - June 2021
 - 18,022 posts
- Survey in private fraud-detection facebook group
 - September 2021, $N = 919$
 - main focus, best data quality: example, classification and confidence of respondent
 - use RAs to classify topics, actions, senders

[back](#)

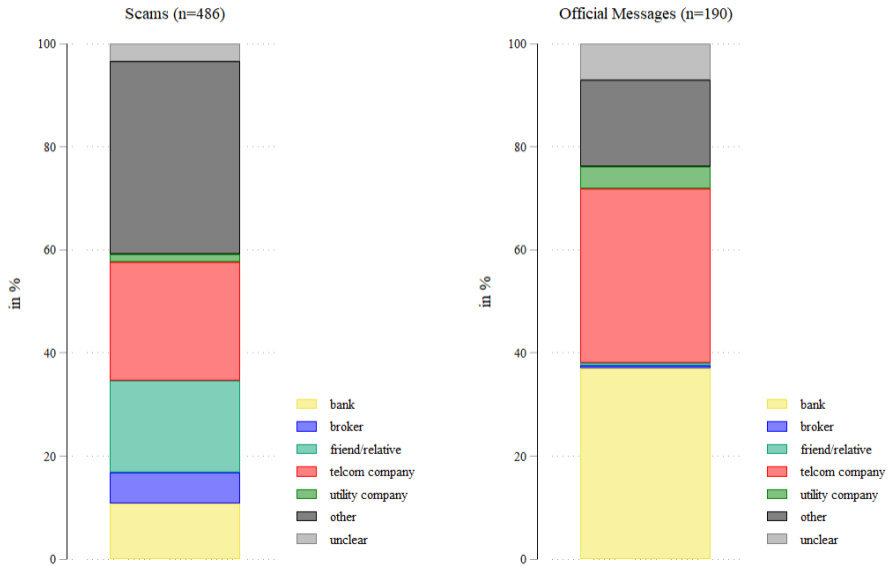
Topics of Submitted Examples



Required Actions of Submitted Examples



(Pretended) Sender of Submitted Examples



No difference for secondary outcomes

Treatment Effects in Block 2

| | Trust in DFS | Response Time SIA | All Scams Identified | All Non-scam Identified |
|-------------------------------|-----------------|-------------------|----------------------|-------------------------|
| Tips (unincentivized) | 0.01 (0.07) | 0.12 (0.08) | 0.10** (0.04) | -0.11*** (0.04) |
| Tips (incentivized) | -0.00 (0.07) | 0.23** (0.10) | 0.11** (0.04) | -0.08* (0.05) |
| Control Mean | 2.02 | 2.21 | 0.30 | 0.52 |
| p-value ($Tips^U = Tips^I$) | 0.92 | 0.26 | 0.99 | 0.42 |
| N | 956 | 956 | 956 | 956 |
| R-Squared | 0.03 | 0.34 | 0.07 | 0.11 |

Notes: The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Scam features

Scam prediction

- off-the-shelf internationally trained model over-predicts in Kenyan data
- Train Electra model to predict which messages are scam
- based on survey and Twitter data, manually labeled by three RAs
- 93% accuracy on test dataset

Scam features

- Scam: locked, win, congratulations, received, hi, hello, interested, kindly
- Official messages: enquiries, queries, dial, transaction, paybill

DFS Use: Balance

| | N | Control Mean | Treatment Mean | p-value |
|--|------|--------------|----------------|---------|
| Any financial transactions on the phone in the past 90 days | 980 | 0.95 | 0.96 | 0.76 |
| Low trust in DFS (0,1) | 1000 | 0.32 | 0.31 | 0.69 |
| Financial transactions ever done with phone: | | | | |
| Sending/receiving funds with mobile money | 1000 | 0.89 | 0.88 | 0.73 |
| Accessing a bank account via your mobile phone | 1000 | 0.50 | 0.52 | 0.49 |
| Paying a bill or paying for something with mobile money | 1000 | 0.72 | 0.71 | 0.94 |
| Taking a mobile loan | 1000 | 0.39 | 0.42 | 0.50 |
| Conducting a financial transaction using an agent (includes withdrawing funds) | 1000 | 0.55 | 0.55 | 0.94 |
| Mobile money used for: | | | | |
| Send money to friends or family | 1000 | 0.82 | 0.83 | 0.72 |
| Receive money | 1000 | 0.79 | 0.79 | 0.88 |
| Receive salary | 1000 | 0.22 | 0.22 | 0.84 |
| Receive payments for business | 1000 | 0.30 | 0.29 | 0.73 |
| Make payments for business | 1000 | 0.39 | 0.38 | 0.77 |
| Pay bills/purchase items | 1000 | 0.74 | 0.74 | 0.77 |
| Save or keep money | 1000 | 0.52 | 0.52 | 0.91 |
| Buy airtime | 1000 | 0.79 | 0.79 | 0.94 |
| Gambling | 1000 | 0.19 | 0.21 | 0.49 |
| Other | 1000 | 0.00 | 0.01 | 0.18 |

Scam Experience: Balance

| | N | Control Mean | Treatment Mean | p-value |
|--|------|--------------|----------------|---------|
| Have you ever been contacted by a scammer? | 999 | 0.96 | 0.97 | 0.41 |
| Last time you were contacted by a scammer: | | | | |
| Less than a week ago | 962 | 0.14 | 0.14 | 0.98 |
| Between 1 week and 4 weeks ago | 962 | 0.22 | 0.23 | 0.78 |
| Between 1 month and 12 months ago | 962 | 0.46 | 0.47 | 0.70 |
| More than 12 months ago | 962 | 0.18 | 0.16 | 0.40 |
| How did you encounter these scams or fraud? | | | | |
| By phone call | 1000 | 0.67 | 0.66 | 0.62 |
| By SMS | 1000 | 0.69 | 0.73 | 0.11 |
| On Whatsapp | 1000 | 0.18 | 0.16 | 0.30 |
| On social media (Facebook, Instagram, ..) | 1000 | 0.20 | 0.19 | 0.85 |
| Other | 1000 | 0.01 | 0.01 | 0.54 |
| What did the scammers ask you to do? | | | | |
| Send money | 1000 | 0.58 | 0.54 | 0.23 |
| Share my password or PIN | 1000 | 0.21 | 0.21 | 0.99 |
| Share my personal information | 1000 | 0.37 | 0.35 | 0.49 |
| Share account details | 1000 | 0.22 | 0.18 | 0.11 |
| Asked for a payment reversal | 1000 | 0.42 | 0.40 | 0.46 |
| Asked to help relative or a friend in need | 1000 | 0.28 | 0.27 | 0.74 |
| Other | 1000 | 0.05 | 0.05 | 0.98 |

Scam Experience (cont'd)

| | N | Control Mean | Treatment Mean | p-value |
|---|------|--------------|----------------|---------|
| How did you know that this was a scam? | | | | |
| Regular number | 1000 | 0.16 | 0.15 | 0.70 |
| From others' experiences | 1000 | 0.48 | 0.43 | 0.13 |
| Requested personal information | 1000 | 0.40 | 0.40 | 0.97 |
| No recent transactions | 1000 | 0.24 | 0.24 | 0.89 |
| Personal awareness | 1000 | 0.47 | 0.44 | 0.34 |
| Incorrectly identified me | 1000 | 0.24 | 0.24 | 0.95 |
| Never used the service | 1000 | 0.23 | 0.20 | 0.19 |
| I did not know the caller/sender | 1000 | 0.34 | 0.36 | 0.40 |
| Unusual time | 1000 | 0.07 | 0.05 | 0.09 |
| Poor language or grammar | 1000 | 0.21 | 0.26 | 0.04 |
| Other | 1000 | 0.02 | 0.01 | 0.23 |
| What did you do? | | | | |
| I fell for it | 964 | 0.62 | 0.65 | 0.48 |
| I ignored it | 964 | 0.08 | 0.07 | 0.46 |
| I deleted it | 964 | 0.26 | 0.24 | 0.67 |
| I reported it | 964 | 0.04 | 0.04 | 0.87 |
| Have you alerted any of your family members or friends? | 961 | 0.90 | 0.88 | 0.30 |
| Have you ever been a victim of a scammer? | 960 | 0.54 | 0.57 | 0.22 |
| Have you alerted any of your family members or friends? | 532 | 0.93 | 0.94 | 0.58 |
| Has anyone you know ever been a victim of a scammer? | 1000 | 0.84 | 0.87 | 0.12 |

Balancing Checks

| | N | Control Mean | Treatment Mean | p-value |
|---|------|--------------|----------------|---------|
| Female | 1000 | 0.51 | 0.49 | 0.57 |
| Age | 1000 | 32.15 | 32.41 | 0.68 |
| Urban (0,1) | 999 | 0.50 | 0.50 | 0.96 |
| Post Secondary Education (0,1) | 1000 | 0.75 | 0.72 | 0.38 |
| Low Income (0,1) | 1000 | 0.77 | 0.79 | 0.36 |
| Formal Employment (0,1) | 997 | 0.37 | 0.36 | 0.94 |
| Internet on Phone (0,1) | 999 | 0.99 | 0.99 | 0.47 |
| Social Media on Phone (0,1) | 999 | 0.99 | 0.99 | 0.75 |
| Shared Phone (0,1) | 1000 | 0.12 | 0.12 | 0.81 |
| Own SIM Card (0,1) | 997 | 0.94 | 0.96 | 0.25 |
| Shared SIM Card (0,1) | 1000 | 0.03 | 0.05 | 0.14 |
| Correctly Identified Messages (Part 1) | 1000 | 4.25 | 4.31 | 0.43 |
| Average Confidence in SIA (Part 1) | 1000 | 4.25 | 4.21 | 0.31 |
| Correctly Identified Scam Messages (Part 1) | 1000 | 2.96 | 2.94 | 0.64 |
| Correctly Identified Official Messages (Part 1) | 1000 | 1.29 | 1.37 | 0.06 |
| Attention Check | 1000 | 0.28 | 0.26 | 0.51 |

Control Variables

| | SIA | | Confidence in SIA | | | | | |
|---------------------|----------------|-----------------|-------------------|-----------------|-------------------|-------------------|-------------------|-------------------|
| Information | 0.13 (0.09) | 0.13 (0.09) | 0.10 (0.10) | 0.11 (0.10) | 0.12*** (0.04) | 0.11*** (0.04) | 0.12*** (0.04) | 0.12*** (0.04) |
| Incentives | 0.11 (0.10) | 0.14 (0.10) | 0.11 (0.10) | 0.12 (0.10) | 0.06 (0.04) | 0.07* (0.04) | 0.07 (0.04) | 0.07* (0.04) |
| Inf. + Inc. | 0.15 (0.09) | 0.18* (0.09) | 0.18* (0.10) | 0.18* (0.10) | 0.08* (0.04) | 0.08* (0.04) | 0.09** (0.05) | 0.09** (0.05) |
| DV in Control | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 |
| N | 1000 | 997 | 937 | 937 | 1000 | 997 | 937 | 937 |
| R-Squared | .014 | .03 | .033 | .039 | .45 | .46 | .45 | .45 |
| Individual Controls | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Experience Controls | | | ✓ | ✓ | | | ✓ | ✓ |
| Design Controls | | | | ✓ | | | | ✓ |

[back](#)

Only those who passed attention check

| | SIA | | | | Confidence in SIA | | | |
|---------------------|--------|--------|--------|--------|-------------------|---------|---------|---------|
| Information | 0.21* | 0.21* | 0.19 | 0.20* | 0.15*** | 0.14*** | 0.15*** | 0.15*** |
| | (0.11) | (0.11) | (0.12) | (0.11) | (0.05) | (0.05) | (0.06) | (0.06) |
| Incentives | 0.16 | 0.19 | 0.16 | 0.16 | 0.07 | 0.08 | 0.09* | 0.09* |
| | (0.12) | (0.12) | (0.12) | (0.12) | (0.05) | (0.05) | (0.05) | (0.05) |
| Inf. + Inc. | 0.14 | 0.16 | 0.19* | 0.19* | 0.05 | 0.05 | 0.07 | 0.07 |
| | (0.11) | (0.11) | (0.11) | (0.11) | (0.05) | (0.05) | (0.05) | (0.05) |
| DV in Control | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 |
| N | 732 | 731 | 685 | 685 | 732 | 731 | 685 | 685 |
| R-Squared | .02 | .043 | .039 | .051 | .44 | .45 | .43 | .44 |
| Individual Controls | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Experience Controls | | | ✓ | ✓ | | | ✓ | ✓ |
| Design Controls | | | | ✓ | | | | ✓ |

[back](#)

Only those who passed attention check

| | Over-Identified | Under-Identified |
|---------------|------------------|--------------------|
| Information | 0.08** (0.03) | -0.09*** (0.03) |
| Incentives | -0.02 (0.04) | -0.03 (0.03) |
| Inf. + Inc. | 0.08** (0.03) | -0.09*** (0.03) |
| DV in Control | .27 | .33 |
| Inf=Inc | .0024 | .021 |
| Inf=Inf+Inc | .98 | .95 |
| Inc=Inf+Inc | .003 | .028 |
| N | 685 | 685 |
| R-Squared | .18 | .11 |

[back](#)