Finding the Fake: Boosting Resistance to Health Misinformation in Jordan with a WhatsApp Chatbot Game

Overview

The code in this replication package constructs the analysis file, tables, and figures for Dugas et al. (2023) from the primary data sources using Stata. One main do file runs all of the code to generate the data, 15 tables, and 11 figures in the paper. The replicator should expect the code to run for about 20-30 minutes.

Data Availability and Provenance Statements

Data Availability

The data for this project are confidential, but may be obtained with Data Use Agreements with Dugas et al. (2023) and the World Bank's Mind, Behavior, and Development (eMBeD) unit. Researchers interested in access to the data may contact eMBeD at eMBeD@worldbank.org. It can take some months to negotiate data use agreements and gain access to the data. The authors will assist with any reasonable replication attempts for two years following publication.

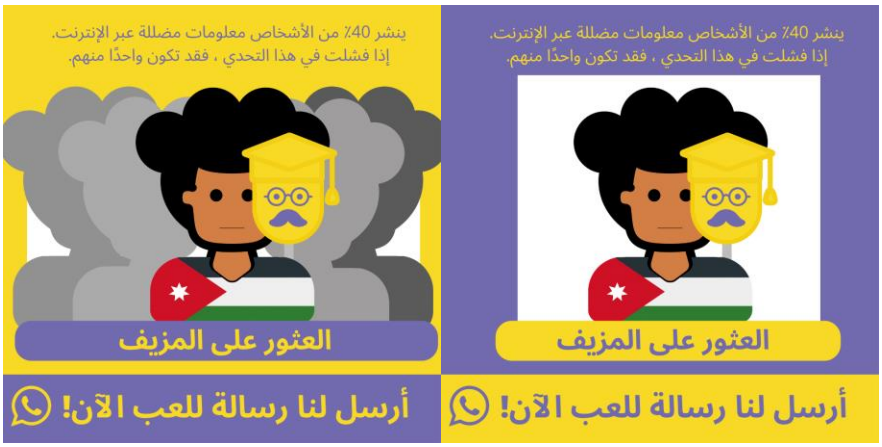
Data Provenance

We conducted a randomized experiment using WhatsApp. This study was approved by the Health Media Lab Institutional Review Board (#2118). We closely coordinated with the Ministry of Health, the Hashemite Kingdom of Jordan during the design, implementation, and analyses of the study.

Recruitment

We conducted this study between October 27, 2022 and November 23, 2022. Participants were recruited through Facebook advertisements targeting users aged 18 years or older and located in Jordan with WhatsApp installed. The advertisements (see Figure 1) marketed our chatbot-based game under the title "Find the Fake", inviting people to play a challenge related to the spread of misinformation online for a chance to win 70 Jordanian Dinar (roughly equivalent to US$ 100).

Figure 1. Facebook Recruitment Ads



After participants clicked the ads, they were automatically directed to a WhatsApp business line, and the chatbot began after participants sent an initial message to the line. In response to the first message sent by participants, the chatbot replied with a message briefly describing the game and participants were asked if they wanted to continue. Participants who opted in were then provided additional background information about the study and contact information for the researchers, completing the informed consent protocol.

Experimental Design

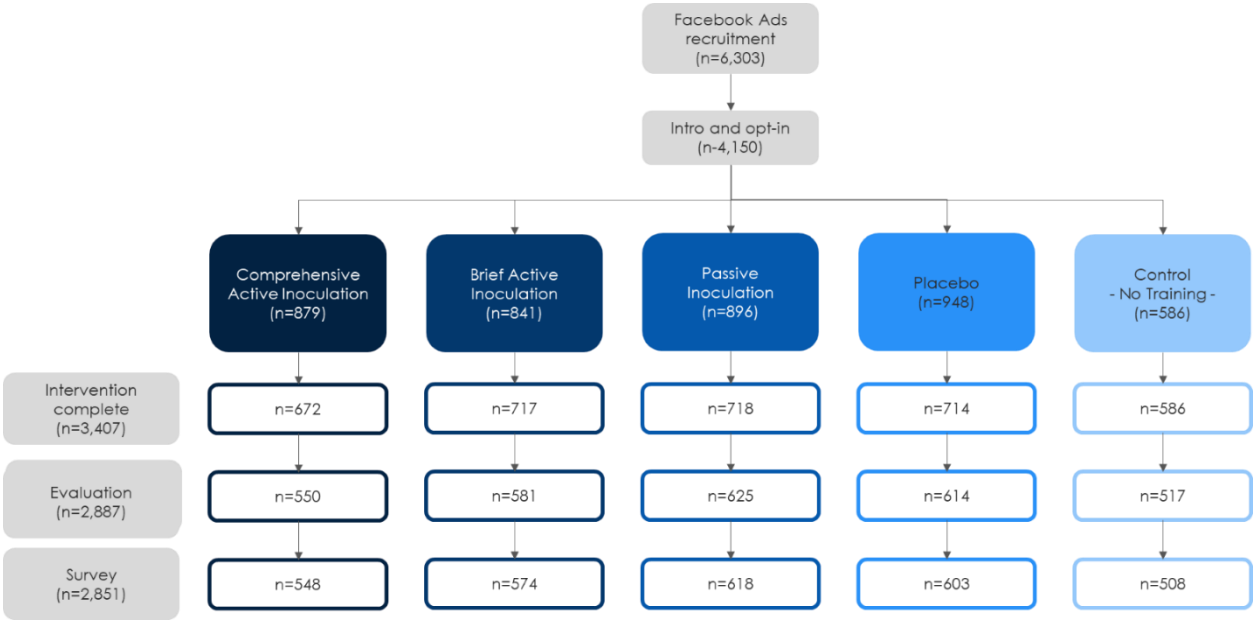
After providing informed consent, participants were randomly assigned to one of five study arms summarized in Table 1.[2] Participants were exposed to educational content, if assigned to a treatment group, then completed a series of questions to evaluate the impact of the intervention, and finally a brief survey assessing attitudes toward the chatbot and collecting basic demographics (see Figure 2 for flowchart).

Table 1. Summary of Study Arms

Study Arm	Description
Comprehensive active inoculation	Completed three lessons on misinformation (appeal to emotions, false experts, digital manipulation of documents/images), including examples of what misinformation looks like and what misinformation does *not* look like. Users received feedback on their performance in the form of scores and badges.
Brief active inoculation	Completed three lessons on misinformation (appeal to emotions, false experts, digital manipulation of documents/images), including only examples of misinformation. Users received feedback on their performance in the form of scores and badges.
Passive inoculation	Exposed to three infographics about misinformation (appeal to emotions, false experts, and how to spot false content), and asked to complete attention checks.

Placebo	Exposed to three infographics on an unrelated topic (road safety), and asked to complete attention checks.
Control	No training or interaction before evaluation.

Figure 2. Flowchart of Experimental Design and Attrition



Comprehensive Active Inoculation

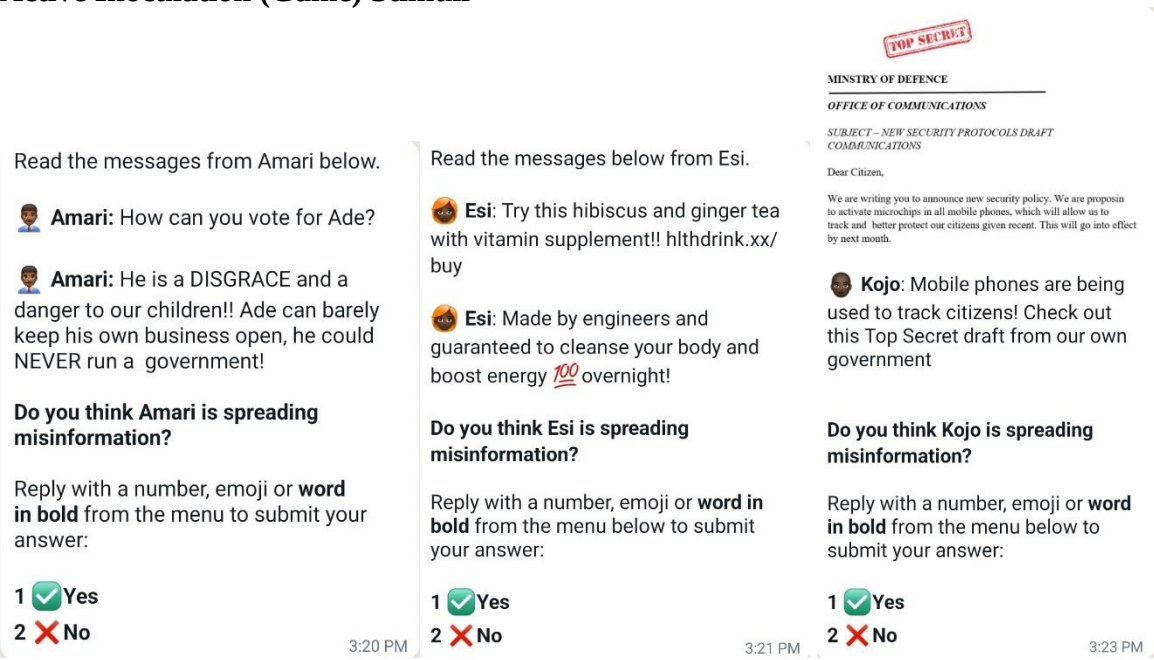
Participants assigned to the comprehensive active inoculation arm completed three rounds of training that addressed three common misinformation techniques: use of content design to elicit strong emotions (e.g., use of very negative words like 'horrible'), reference to false experts (e.g., appealing to authorities who are expert in a domain unrelated to the issue at hand), and the digital manipulation of documents/images with fake content.

In each round, participants were presented with different vignette-based scenarios reflecting domains where misinformation is commonly spread including political and health issues. After a brief description of the scenario, participants were presented with messages shared by different characters and asked to indicate whether they believed the messages were spreading misinformation or not. In the comprehensive version of the game, messages included examples using misinformation spreading techniques and examples of messages that did *not* use misinformation techniques.

After each question, participants would get feedback about whether their responses were correct or incorrect, and the chatbot leveraged several gamification strategies. For example, participants received visual feedback about their scores that persisted throughout each round (e.g., stars for correct responses and blank squares for each incorrect response). In addition, participants who correctly identified the message spreading

misinformation in a round received a virtual 'badge'. At the end of the game, participants received a recap of their performance and the various badges they earned. See screenshots from the game below.

Active Inoculation (Game) Stimuli



Brief Active Inoculation

The brief active inoculation condition was the same as the comprehensive inoculation game except participants were exposed to only examples of messages that used misinformation techniques. To shorten the game, all neutral examples were excluded from the training rounds.

Passive Inoculation

Participants assigned to passive inoculation were sent three infographics adapted from the UNESCOs graphics on media and information literacy (UNESCO, 2022) with imagery and text referencing COVID-19 removed. The three infographics (see Appendix B) covered information that paralleled the lessons taught in the three rounds of gamified training. Specifically, they addressed: (1) that disinformation was often designed to trigger emotional reactions, (2) that false experts are often used to spread manipulative information, and (3) how to spot false content and rumors. The infographics were described to users as tips that would help them perform well on the 'game' that began after they learned the tips, which actually consisted of our evaluation questions. After each infographic, participants were asked to answer an attention check question (e.g., 'Please reply with "3". Do not reply with other options.'). The attention check served as a placebo for interaction similar to the game experience, but without actively testing users' understanding of the misinformation lesson.

Placebo
The procedure of the placebo arm followed the same protocol as the infographics study arm except all infographics were unrelated to misinformation. The featured infographics were published by the World Health Organization (WHO, 2022) and instead focused on three topics of road safety: (1) traffic injury facts; (2) speed management; (3) how to keep children safe (see Appendix C for placebo infographics). This activity was framed as a practice round for users in which they could learn the game mechanics of how to respond to questions, with the 'real' game (i.e., evaluation) beginning after the practice.

Control
In the control condition, there was no exposure to any material, and participants proceeded directly to the outcome evaluation. The outcome evaluation was framed as the game for participants in this study arm.

Measures
Evaluation Outcomes. To evaluate the impact of the misinformation treatments, the primary outcomes of interest were: (i) rates of accurately discerning between headlines that use misinformation tactics and those that don't, and (ii) discernment in sharing headlines (see Appendix E for evaluation measures).

Misleading Headlines. All headlines focused on COVID-19 as a theme, and we generated new headlines based on existing headlines rather than using headlines from published articles. This approach was adopted to ensure that all headlines were completely novel to participants whereas participants could vary in their familiarity with real-world headlines.

The designed headlines were adapted from prior research testing the effectiveness of inoculation against COVID-19 misinformation (Basol et al., 2021), and reflect the three common tactics used in misinformation that were also the focus of the inoculation training (i.e., emotional appeals, false experts, and fake documents/images). Additionally, the misleading headlines were designed to address themes similar to misinformation that had spread online according to various fact-checking sources (e.g., AFP Fact Check, Africa Check) to enhance ecological validity.

Judgements of Misinformation. Adapted from Roozenbeek et al (2022), participants were presented with six headlines, three using misinformation tactics and three that did not use common misinformation tactics. When presented with each headline, participants were asked to respond to the question '*Does this headline use any misinformation techniques*?' on a 4-point scale: Definitely is misinformation, Probably is misinformation, Probably is not misinformation, Definitely is not misinformation. For ease of interpretation, ratings were scored such that higher scores represent stronger belief that a headline was misinformation (*Definitely is misinformation* = 4 and *Definitely is not misinformation* = 1).

In line with prior literature, we compute three scores to assess accuracy in misinformation detection (Basol et al., 2021; Maertens et al., 2021). First, we calculate a measure of *discernment*, defined as a participant's average misinformation scores for misleading

headlines minus their average score for headlines without misleading content. With this operationalization, discernment scores could range from -3 to +3 where a score of +3 indicates a participant rated all misinformation headlines as 'Definitely misinformation' and all non-misinformation headlines as 'Definitely not misinformation' for perfect discernment.

We also examine the disaggregated discernment score including the average ratings for the three misleading headlines and ratings for the three non-misleading headlines. As higher ratings correspond to judgements that a headline is using misinformation tactics, more accurate scores would be represented by higher scores on the misleading headlines (representing true positives) and lower scores on the non-misleading headlines (representing true negatives).

Sharing Misinformation*.* Adapted from Basol et al. (2021) and Roozenbeek and van der Linden (2020), sharing of misinformation was assessed with two headlines: one that did not use misinformation tactics and a headline that used misinformation tactics. Each participant was randomly assigned to one of two misleading headlines, a headline that used extreme emotion or a headline that used a false expert. Participants were asked to rate their likelihood of sharing each of the headlines on a four-point scale: (1) Very unlikely to share, (2) Unlikely to share, (3) Likely to share, (4) Very likely to share. As with judgements of misinformation, we report three scores for sharing—discernment of sharing, likelihood of sharing misleading headlines, and likelihood of sharing non-misleading headlines.

Self-Report Outcomes. Complementing the evaluation of the intervention's impact on detection and sharing of misinformation, we examined differences in attitudes toward the chatbot game related to perceived impact and engagement.

***Improved Confidence*.** Participants were asked to indicate the extent to which they felt more, the same, or less confident in detecting misinformation after completing the game.

***Perceived Difficulty*.** Participants were asked to report whether they thought the game was too difficult, the right level of difficulty, or too easy.

***Recommending the Game*.** Participants were asked to indicate the extent to which they would recommend the game to others with three possible response categories: yes, maybe, and no.

Participants

A total of 2,851 participants completed the study. Of them, 63% identified as male and 33% as female; 49% reported having completed secondary education (3% with no education, 12% with primary, and 3% with tertiary education); 53% reported being between the ages of 18 and 29, 25% in their 30s, 13% in their 40s, and 5% over the age of 50. Finally, 85% reported being vaccinated for COVID-19, 5% unvaccinated but willing to vaccinate, and 5% unvaccinated and unwilling to vaccinate. None of the treatment arms reported significant differences in demographics and vaccination status compared to the control group. See Appendix F for sample composition and randomization balance.

Statement about Rights

We certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

No data can be made publicly available.

Details on Data Source

Data.Name	Data.Files	Location	Provided	Citation
"Experiment data"	reproducibility_data.dta		TRUE	Dugas et al. (2023)

Computational requirements

Software

Stata (code was last run with version 18)

- `Missings` (as of 2023-12-13)
- `Outreg2` (as of 2023-12-13)
- `wyoung` (as of 2023-12-13)
- `coefplot` (as of 2023-12-13)
- `grc1reg` (as of 2023-12-13)
- `violinplot` (as of 2023-12-13)
- `dstat` (as of 2023-12-13)
- `moremata` (as of 2023-12-13)

- The initial code of "reproducibility_do.do" at the beginning will install all dependencies locally, and should be run once.

Memory and Runtime Requirements

Summary

Approximate time needed to reproduce the analyses on a standard (2023) desktop machine:

- Over 60 minutes

Details

The code was last run on an **Intel-based laptop with Windows 10, 16.0 GB RAM.**
Computation took about 1 hour.

Description of programs/code

- Programs in `reproducibility.do` will run all analyses and generate all tables and figures in the main body of the article and the appendix. The file `programs/reproducibility_do.do` will run them all. Each program called from `main.do` identifies the table or figure it creates. Output files are called appropriate names (`table5.tex`, `figure12.png`) and should be easy to correlate with the manuscript.

License for Code

The code is licensed under an MIT license. See the "license.rtf".

Instructions to Replicators

- Edit "`global dir`" to adjust the default path
- Run "reproducibility_do`.do`" once on a new system to set up the working environment and to run all steps in sequence. Download the confidential data files referenced above in the same default folder along with the do file. Each should be stored in the format that you download them. No further action is needed on the replicator's part.

Details

- reproducibility_do`.do`: will create all output (figures and tables).
  - These programs were last run at various times in 2023.

List of tables and programs

The provided code reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper

Figure/ Table #	Program	Line Number	Output file	Note
	Reproducibility_do.do	345	Table X.xls	All requires confidential data
Figure 2		295		
Table 2		324	Table 2.xls	
Table 3		353	Table 3.xlsx	
Table 4		378	Table 4A.xls, Table 4B.xls	
Figure 3		406	Figure 3A.png Figure 3B.png	
Table 5		430	Table 5.xlsx	
Table 6		452	Table 6A.xls Table 6B.xls	
Figure 5		531	Figure 5.png	
Appendix E		555	Appendix E.xlsx	
Appendix G		674	Appendix G.xlsx	
Appendix H		716	Appendix H1.png Appendix H2.png Appendix H3.png Appendix H4.png Appendix H5.png	

		Appendix H6.png
Appendix I	739	Appendix I1.png
		Appendix I2.png
Appendix J	769	Appendix J1.xls
		Appendix J2.xls
		Appendix J3.xls
		Appendix J4.xls
		Appendix J5.xls
Appendix K	913	Appendix K.xls

Acknowledgements

Some content on this page was copied and adapted from Dugas et al. (2023) with the authors' permission.