

## RESEARCH ARTICLE

# Inoculating against extremist persuasion techniques – Results from a randomised controlled trial in post-conflict areas in Iraq

Nabil Saleh<sup>1</sup>, Fadi Makki<sup>1,2</sup>, Sander Van Der Linden<sup>3</sup>, and Jon Roozenbeek<sup>3,\*</sup>

Received: December 15, 2022 | Accepted: September 03, 2023 | Published: September 11, 2023 | Edited by: Jonas R. Kunst

1. Nudge Lebanon; 2. Boston Consulting Group; and 3. Department of Psychology, University of Cambridge.

\*Please address correspondence to Jon Roozenbeek, [jjr51@cam.ac.uk](mailto:jjr51@cam.ac.uk), Department of Psychology, University of Cambridge.

This article is published under the Creative Commons BY 4.0 license. Users are allowed to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Extremist organisations often use psychological manipulation techniques to persuade new members to join. Previous research has found that people can be made more aware of such techniques through psychological “inoculation” interventions, which seek to foster resistance against unwanted persuasion attempts. We conducted a field experiment ( $N = 191$ ) in post-conflict regions of Iraq to assess the effectiveness of a short inoculation game, *Radicalise*, in improving vulnerable individuals’ resistance against extremist manipulation techniques. In a conceptual replication of Saleh et al. (2021), we translated and adapted the game for the Iraqi context and then conducted a 2x2 mixed (pre-post / treatment-control) randomised controlled experiment among a group of vulnerable youth in areas previously under ISIS control. We included two outcome measures: participants’ ability and confidence in correctly assessing WhatsApp messages making use of extremist manipulation techniques, and the ability to identify the factors that make an individual vulnerable to extremist recruitment. We find that playing the game significantly improved participants’ ability ( $p = 0.034$ ,  $d = 0.31$ ) in spotting manipulative messaging while the improvement in participants’ confidence fell just above the traditional 0.05 significance level ( $p = 0.051$ ,  $d = 0.29$ ). However, unlike in Saleh et al. (2021), we find that playing the game did not impact participants’ ability to identify vulnerable individuals ( $p = 0.896$ ,  $d = 0.02$ ). We note that our field study may have been underpowered compared to the original study and our results should therefore be interpreted with some appropriate caution.

**Keywords:** Inoculation Theory; Prevention of Violent Extremism; Field Research; Post-conflict Region; Gamification

## 1. INTRODUCTION

The internet provides a fertile environment for extremist organisations seeking to recruit new members (Knudsen, 2018). For example, the Islamic State of Iraq and Syria (ISIS) saw the internet as a crucial tool to disseminate propaganda and present an idealistic view of life in the so-called “Islamic State” (Berger & Morgan, 2015). People who are vulnerable to such recruitment attempts often come from marginalised communities, or are otherwise disadvantaged (Doosje et al., 2016). Aside from emphasising (and claiming to provide a solution to) socio-economic disadvantages such as poverty and marginalisation, extremist recruiters are also known to exploit potential recruits’ psychological vulnerabilities (Hogg et al., 2013; Kruglanski et al., 2014). Strategies to exploit such vulnerabilities include identifying target individuals who may be amenable to radicalisation, gaining their trust, isolating them from their environment (e.g., friends and family), and peer-pressuring them into committing acts of violence (Doosje et al., 2016; Kruglanski et al., 2014).

Iraq is one of the world’s foremost countries where terrorism poses a significant problem (Institute for Economics & Peace, 2022). Following the US-led invasion of Iraq in 2003, the country faced cycles of civil unrest and violence that culminated with the rise of what is known as the Islamic State (IS), which at its peak controlled about 40% of Iraqi territory (BBC News, 2016; Glenn et al., 2019). Five years after victory against the IS was announced by Iraqi security forces in the Battle for Mosul in July 2017, Iraq watches warily as remnants of the terrorist group continue to conduct sporadic attacks and operate within desert areas not far from the Syrian border (Kittleson, 2020; Kouachi, 2022; Starr & Tawfeeq, 2020). The constant threat of IS and other extremist groups poses a risk to vulnerable populations in under-served areas in the country, particularly among youth.

According to the World Bank, Iraq is experiencing a “youth bulge” (Amirali, 2019), having one of the youngest populations in the world with a numerical dominance of young men aged 15-29 relative to other age groups in society, with roughly 60 percent of the country’s population under 25 years of age (World Bank, 2020). To add to that, social media usage is on the rise, with estimates suggesting that 67.2 percent of the “eligible” population (i.e., 13 years old and above) in Iraq used Facebook in 2022 (Kemp, 2022). Recent studies indicate that youth literacy levels are low: 33% of youth between the ages of 15 and 29 are illiterate or only semi-literate, 33% have completed primary school, 28% have finished middle or high school, and only 7% have completed post-secondary education (Amirali, 2019). In addition, Iraqi youth are experiencing widespread unemployment combined with high levels of political exclusion, sectarian politics, militarisation, perceptions of injustice, frustrated aspirations, war-related trauma, and the rapid breakdown and transformation of traditional institutions such as family and tribe (Amirali, 2019).

Such factors are often labelled as drivers that could increase vulnerability to recruitment attempts by extremist organisations (UNODC, 2018). For example, some studies show that continued exposure to violence can lead to desensitisation, which is defined as diminished emotional responsiveness to violence (Mrug et al., 2016). Because of this, young people in post-conflict areas of Iraq might be more desensitised to the rhetorical strategies used by recruiters working for extremist organisations. This, in turn, may mean that baseline perceptions of manipulative messaging strategies are different in Iraq than in many Western countries. This is one of the questions we investigate in this study.

### 1.1 Prebunking and Inoculation Theory

There are various methods available to counter misinformation and associated manipulation

attempts, one of which is pre-emptive debunking or prebunking (Roozenbeek et al., 2023). There are numerous approaches and methods that fall under the “prebunking” umbrella, which can be as simple as providing a pre-emptive correction or a generic warning about misinformation (Ecker et al., 2022).

Perhaps the most influential form of “prebunking” interventions draws on inoculation theory (Compton et al., 2021). Pioneered by social psychologist William J. McGuire in the 1960s, inoculation theory posits that people can develop psychological resistance against unwanted persuasion attempts (McGuire, 1964) much like people gain resistance to viruses via medical vaccines. Inspired by a biomedical analogy, psychological inoculations involve pre-emptive exposure to a “weakened dose” of the persuasive argument, which subsequently reduces susceptibility to the “real” persuasion attempt (Eagly & Chaiken, 1993; McGuire & Papageorgis, 1961, 1962). Meta-analyses and reviews of the field have shown that inoculation can robustly strengthen people’s attitudes against future persuasion and manipulation attacks (Banas & Rains, 2010; Compton et al., 2021; Lewandowsky & Van Der Linden, 2021; Lu et al., 2023; Traberg et al., 2022).

Psychologists have proposed that the underlying psychological mechanisms of inoculation interventions involve both *threat* and *pre-emptive refutation* (Compton, 2021; McGuire & Papageorgis, 1961, 1962). Threat is described by Compton et al. (2021) as “the motivational force that triggers such responses as counter-arguing against future challenges to a position” (p.1). The threat component of inoculation messages involves making people aware that an attack on their beliefs is imminent, for example by forewarning them that insincere actors might be out to manipulate them or influence their beliefs (Linden et al., 2017). Doing so provides the motivation to subsequently resist persuasion. The pre-emptive refutation component involves providing inoculated individuals with the

cognitive tools to resist persuasion, by giving them concrete information or skills that they can use to generate counterarguments, and strengthen their attitudes against persuasive attacks, such as information on how to spot specific manipulation techniques (Traberg et al., 2022). In a recent study, Maertens et al. (2023) further explore the underlying mechanisms of various types of inoculation interventions, and propose a “memory-motivation” model that underlies their long-term effectiveness. In other words, while memory (i.e., learning) plays a leading role in explaining how inoculation interventions build resistance against misinformation and other forms of manipulation, motivation and (motivational) threat are secondary but nonetheless likely important mechanisms (e.g., see Basol et al. (2021).

In recent years, researchers have explored whether psychological resistance can be conferred not only against individual misleading arguments, but against the manipulation techniques that underlie misinformation in general, such as conspiratorial reasoning, emotional manipulation, or logical fallacies (Traberg et al., 2022). This approach, called *technique-based inoculation* (Roozenbeek et al., 2023), significantly improved people’s ability to identify manipulative online content, including in real-world social media environments such as YouTube where exposure to misinformation is common (Roozenbeek, Traberg, & Van Der Linden, 2022).

One common form of technique-based inoculation interventions are online games and quizzes such as *Bad News* (Roozenbeek & Van Der Linden, 2019), *Cranky Uncle* (Cook et al., 2022) and “*Spot the Troll*” (Lees et al., 2023), which—through simulated scenarios—expose players to weakened doses of common misinformation tactics along with ways on how to spot them. These interventions forewarn players that the manipulation techniques they will learn about in the game can be used to mislead them (the *threat* component of the in-

oculation), and pre-emptively refute these techniques through perspective-taking exercises and providing narrative feedback on players' performance (Roozenbeek & Van Der Linden, 2020).

Such game-based interventions are often referred to as a form of "active" inoculation (Mcguire & Papageorgis, 1961). Active inoculation interventions rely on interactive skill development as players are encouraged to generate their own content and think about their decisions, which helps trigger the creation of counterarguments against the misinformation that they receive a weakened dose of within the game environment (Traberg et al., 2022). Conversely, "passive" inoculation involves simply providing people with an inoculation message without feedback (Green et al., 2022; Roozenbeek, Van Der Linden, et al., 2022). Research has shown that these games can confer psychological resistance against many kinds of manipulation techniques that commonly underlie misinformation, such as about climate change, COVID-19, and politics (Cook et al., 2022; Traberg et al., 2022; Van Der Linden & Roozenbeek, 2020).

Although inoculation messages have been found to be superior to both supportive messages (promoting an attitude already held) and no-treatment controls at conferring psychological resistance (Banas & Rains, 2010), there are also some limitations to the theory. Its effectiveness can depend on various factors, such as the complexity of the issue being addressed and the receptiveness of the audience. For example, some individuals may be more resistant to persuasion than others, due to factors such as personality, ideology, or prior experiences (Traberg et al., 2022). Additionally, although prebunking has shown to be superior to debunking conspiracy theories (O'mahony et al., 2023), some issues may be so complex or emotionally charged that inoculation may not be sufficient to change deeply held beliefs or attitudes (Compton et al., 2016). Another

limitation is the sustainability of the developed resistance which starts to fade after few days, weeks, or months of being exposed to the inoculation intervention (Maertens et al., 2021). However, research has shown that the longevity of inoculation effect can be extended through the application of 'booster shots' (Maertens et al., 2023), and can generate resistance to persuasion even after multiple persuasive attacks (Ivanov et al., 2018). Finally, although psychological inoculation runs on a spectrum from prophylactic to more "therapeutic" applications depending on prior exposure (Compton et al., 2021), it is often assumed to be more effective when done pre-emptively so we focus on youth in this study who are at-risk but have not yet been radicalized.

## 1.2 The Present Study: The *Radicalise* and *MindFort* Games

In 2020, the authors applied insights from the literature on gamification and technique-based inoculation to develop *Radicalise*, a 10-minute browser game aimed at improving people's ability to recognise manipulation and persuasion techniques commonly used by extremist organisations to recruit new members (Saleh et al., 2021). These techniques were taken from the broader literature on extremism and psychological manipulation: identifying vulnerable individuals (Bartlett et al., 2010; Knudsen, 2018; Precht, 2007); gaining their trust (Doosje et al., 2016; Walters et al., 2013); isolating them from their community (Doosje et al., 2016; Ozer & Bertelsen, 2018); and pressuring them to commit an act (usually of violence) in the name of the extremist organisation (Doosje et al., 2016; Ozer & Bertelsen, 2018; Precht, 2007). Table 1 shows a summary of the four levels in the *Radicalise* game, and the manipulation technique that each level represents.

By exposing the player to weakened doses of the techniques that are used by extremist recruiters and by demystifying them in a simulated setting, the aim of *Radicalise* is to

**Table 1**The four levels in the *Radicalise* game.

| Level                             | Details  |
|-----------------------------------|--|
| <b>Level 1:</b><br>Identification | The first level of the game introduces the player to the game environment, the radical organisation they 'work' for and its goals, and their role in this organisation to recruit individuals for its cause. The level then proceeds to deliver on two key lessons: where and more importantly whom to recruit, focusing on the reasons commonly cited in literature leading to people becoming radicalised. Some of these reasons are related to motivational dynamics including a quest for identity, a search for purpose, and personal significance, the pursuit of adventure, or circumstantial and environmental factors such as extended unemployment or disconnection from society due to incarceration, studying abroad, or isolation/marginalisation (Bartlett et al., 2010; Borárosová et al., 2017; Doosje et al., 2016; Kruglanski et al., 2014; Thornton & Bouhana, 2017). |
| <b>Level 2:</b><br>Gaining Trust  | Following the identification level, the second level instructs players to start reaching out to their target for the purpose of gaining their trust through offering them acceptance and the promise of friendship. The objective of this level is to teach players how radical groups target vulnerable individuals when they feel uncertain by providing them with a sense of an in-group, motivating them to identify with the group and reducing their uncertainty by setting clear norms and values from this group (Doosje et al., 2016; Euer et al., 2014; Hogg et al., 2013).  |
| <b>Level 3:</b><br>Isolation      | Once the target's trust has been gained, players must pursue the indoctrination of their target into the group's ideology. The player is instructed to isolate the target from their old social environment by encouraging them to burn bridges with family and friends who do not belong to the group. In the meantime, the player is asked to continue working on making their new in-group stronger and more cohesive. The game also introduces the player to a number of "catalysts" that enforce quicker physical and psychological isolation for the target through bonding with the organisation, culminating by asking the target to travel overseas to stay with at the premises of the organisation (Doosje et al., 2016; Precht, 2007; Stein, 2017; Swann et al., 2012).  |
| <b>Level 4:</b><br>Activation     | In the fourth and final level, the grooming process has gotten the target to a point of near perfect alignment with the organisation's ideology. In this level, the player experiences putting "the last nail in the coffin", getting the player to make the target commit an act of violence as an ' <i>initiation ritual</i> ' through exercising pressure and using a certain amount of coercion on the target (Doosje et al., 2016).   |

help people identify and resist these manipulation techniques when they encounter them online. The game served as a simulation to help familiarise players with these tactics, as perceived vulnerability and familiarity are key to being able to identify unwanted persuasion attempts (Sagarin et al., 2002). An effectiveness study conducted in the UK suggests that playing *Radicalise* yields a significant improvement in participants' ability and confidence in spotting manipulative messages that could potentially be used by extremist groups as well

as an improvement in their ability to identify the characteristics associated with psychological vulnerability to radicalisation (Saleh et al., 2021).

The current study had two goals: first, to replicate the original research by Saleh et al. (2021) conceptually but also to examine the robustness of the intervention in Iraq's specific conditions. Second, the study aimed to evaluate how the game impacted individuals from conflict and post-conflict regions. As discussed in the

Introduction section, conducting research in post-conflict areas of Iraq, where many factors are present that may increase vulnerability to extremist recruitment and manipulation attempts, not only presents a challenge in terms of replicating prior work (for instance in terms of participant recruitment) but also serves as an extension of the effectiveness of inoculation interventions outside of the Global North.

For this study, we collaborated with Spirit of Soccer ([www.spiritofsoccer.com](http://www.spiritofsoccer.com)), which runs football clinics in post-conflict areas throughout Iraq to engage beneficiaries through soccer-based activities that help to promote important life skills such as communication, teamwork, and conflict resolution as well as provide mine risk education in areas affected by landmines or other explosive remnants of war.

As such, we conducted a field experiment in areas where such a tool is most needed: post-conflict regions in Iraq, namely Mosul and Duhok, which were fully or partially occupied by ISIS. Our first step was to translate the *Radicalise* game's script and user interface from English to Arabic. To account for the low literacy level for the youth group with whom the study was to be conducted, the script was written phonetically using local Iraqi Arabic. The translation was carried out by native speakers of Iraqi Arabic and was reviewed through several iterations to ensure that the translation closely conveys the essence of the English version, which used a type of humour that is associated with Western popular culture. This required rewriting part of the script to employ locally relevant humour and other context-related materials. The second measure was to cleanse the script from any content that may be perceived as harmful, offensive, or otherwise damaging given the volatile context, security concerns, and political sensitivities in the areas in which we conducted our study. Accordingly, any phrases or words that relate to extremism, violence, or conflict had to be altered to have a

more neutral connotation, including changing the name of the game, from *Radicalise* (English version) to *MindFort* (Arabic version). Finally, as in the English version of the game, all materials were ensured to be areligious, apolitical, and neutral. See Figure 1 for screenshots from the game.

The present study departs from the original study, Saleh et al. (2021), in several important ways. Aside from the obvious differences in terms of sample, context, and language, we also changed the name of the game from *Radicalise* to *MindFort*. Additionally, there were differences in recruitment strategies (an online panel versus field recruitment), and even seemingly minor and culturally sensitive differences in the wording of materials could have significant impact on participants' responses. These differences are discussed further in the Discussion section. Acknowledging these differences, we designed the study to test the following hypotheses (Saleh et al., 2021)<sup>1</sup>:

**H1:** After playing *MindFort*, people perform significantly better at identifying manipulation techniques used in extremist recruitment compared to a control group.

**H2:** *MindFort* players perform significantly better at identifying characteristics that make one vulnerable to extremist recruitment compared to a control group.

**H3:** *MindFort* players become significantly more confident in their assessments.

## 2. METHODS

To test the effectiveness of the *MindFort* game as a method of psychological inoculation, we conceptually replicated the initial *Radicalise* study (Saleh et al., 2021). Specifically, we conducted a 2x2 (pre-post / treatment-control) randomised controlled trial in Mosul and Duhok in February of 2022. The survey was administered on site on tablets and mobile phones, using the survey software Qualtrics. The study design,

## Figure 1

Screenshots of the *MindFort* Landing Page (left) and Game Environment (right).



measures, hypotheses, and statistical analyses are the same as original study. Data, Qualtrics files, analysis and visualisation scripts, and the supplementary information can be found in our OSF repository (Saleh et al., 2023). This study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2021.024).

### 2.1 Participants

Our sample consisted of a group of vulnerable youth from Mosul and Duhok between the ages of 18 and 40. This group was identified by our implementing partners, Spirit of Soccer. To identify its beneficiaries, Spirit of Soccer partners with local organisations and community leaders who have knowledge of the communities and can help to identify young people who are most at risk, including those who live in areas affected by conflict, have limited access to education or other opportunities, or who have experienced trauma or loss due to conflict. In addition, Spirit of Soccer conducts its own needs assessments and surveys to better understand the needs and vulnerabilities of the communities it serves. Accordingly, it tailors its various programmes and activities, including peacebuilding workshops, youth leadership development programs, soccer clinic sessions, and mine risk education programs, to effec-

tively address the specific needs of its beneficiaries.

In the original study, the sample size was calculated by conducting an *a priori* power analysis using G\*Power,  $\alpha = 0.05$ ,  $f = 0.26$  ( $d = 0.52$ ) and a power of 0.90, with two experimental conditions, based on prior research with similar designs (Basol et al., 2021, 2020; Van Der Linden & Roozenbeek, 2020). Accordingly, the minimal sample size required for detecting the main effect was approximately 156 (78 per condition). While the original study's preregistration indicated 260 individuals, the final sample was slightly higher with 291 individuals.

For this study, we sought to recruit as many participants as possible, hoping to achieve a similar sample size as the original study ( $N = 291$ ). This proved to not be feasible due to several constraints including difficulties associated with conducting field studies, particularly in high-risk areas, as well as limited access to participants, budget limitations, and time constraints. In total, we collected 221 responses. We used the following exclusion criteria: 1) incomplete response, 2) no informed consent provided, 3) incorrect completion code (inoculation group; participants who played the *MindFort* game received a code at the end of the final level which they needed to proceed with the rest of

the survey). After excluding participants using these criteria, the final sample was reduced to 191 individuals. A sensitivity analysis using G\*Power software (sensitivity, ANCOVA,  $\alpha = 0.05$ ,  $N = 191$ , Numerator  $df = 1$ ) shows that the smallest effect-size we can reliably detect is  $f = 0.20$ . This indicates that the study may be underpowered, meaning that the sample size may not have been sufficient to detect significant differences given the observed effect sizes.

The final sample consisted of 71.7% men (28.3% women), with a mean age of 25.8 years ( $SD = 6.88$ ). Most participants reported being single (50.8%) or married (44.5%), with a minority being divorced (4.2%) or separated (0.5%). 63.9% had previously participated in a different Spirit of Soccer workshop. The sample of 191 was randomly assigned to control and treatment conditions, ( $n = 96$  and  $n = 95$ , respectively). The randomisation ensured that the four demographic variables did not differ by condition. For gender, of the 96 individuals in the control group, 31 were female and 65 were male, while of the 95 individuals in the treatment group, 23 were female and 72 were male. The results of the analysis revealed that there was no statistically significant association between treatment and gender, ( $\chi^2 = 1.54$ ,  $p = 0.215$ ). For age, the mean score of control condition was 25.08 ( $SE = 0.62$ ) and the mean score of treatment condition was 26.59 ( $SE = 0.78$ ). A  $t$ -test showed that the difference in means was not statistically significant, ( $t(189) = -1.52$ ,  $p = 0.131$ ). For marital status, the results of chi-square test revealed that there was no statistically significant association between treatment and marital status ( $\chi^2 = 1.60$ ,  $p = 0.660$ ). Finally, for previous participation in SoS workshops, a chi-square test showed that there was no statistically significant association between treatment and participation in the SoS program ( $\chi^2 = 0.16$ ,  $p = 0.691$ ). See Table S1 in the Supplementary Information for the full sample descriptives.

## 2.2 Procedure

Like in the original study, participants were first shown a series of six WhatsApp posts containing conversations in which one individual made use of one of three manipulation techniques learned in the *MindFort* game (gaining a target's trust, attempting to isolate them from their environment, and prompting them to commit an act of violence), see Figure 2. After every WhatsApp post, participants were then presented with questions about how manipulative they perceived each post to be, and how confident they were in this assessment – more on this in the Measures section.

Next, participants were shown a series of 8 vignettes that described the profile of either an individual who displayed the characteristics of someone who may be vulnerable to being recruited by an extremist organisation (6 vignettes) or someone who does not display such characteristics and is thus not more likely to be vulnerable to being recruited (2 vignettes). See Figure 3. After every vignette, participants were presented with questions about the likelihood of this person joining a group with different beliefs & confidence – more on this in the Measures section.

Participants were then randomly assigned into a treatment or control group ( $n_{MindFort} = 95$ ,  $n_{control} = 96$ ). The treatment group participants played the *MindFort* game, while control group participants were asked to play *Tetris* for a similar period to what it takes to finish *MindFort*. After this, participants again rated the same WhatsApp conversations and vignettes from the pre-test.

## 2.3 Measures

After being presented with each of the WhatsApp conversations, participants were then asked the following questions (on a 1-7 Likert scale ranging from 1 “not at all” to 7 “very”): “How manipulative do you find the sender’s messages?” and “How confident are you in



### Figure 2

**Example of a WhatsApp Message Stimulus.** Translation, Message 1: “May peace be upon you, brother. How are you doing? How is everything? I hope all is well”. Message 2: “May peace be upon you too. All is well here. What about you? I hope all is well on your side too”. Message 3: “Great. I wanted to let you know that I was contacted by Abu Ibrahim. He is looking for a new driver. Are you ready?”. Message 4: “I apologise. Unfortunately, I am not ready for this job. Please forgive me”. Message 5: “This is nonsense! Listen...If you do not cooperate, we will expose you to your family and let them know the kind of person you are.”

| كثيراً                | 6                     | 5                     | معتدل                 | 4                     | 3                     | 2                     | على الإطلاق           |  |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | ما مدى فلتاعب التي تجده في رسائل المرسل؟ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | ما مدى تكلف في جوابك؟                    |

### Figure 3

**Example of a Profile Vignette.** Translation: “Yasin worked in the IT sector after graduating from university. Yasin stopped working after his company declared bankruptcy more than 9 months ago and, recently, he stopped looking for a job after many attempts. Yasin spends his time now at home on his computer”.

كان ياسين يعمل في مجال الكمبيوتر بعد تخرجه من الكلية. توقف ياسين عن العمل بعد إفلاس الشركة منذ أكثر من 9 أشهر وتوقف مؤخرًا عن البحث عن وظيفة جديدة بعد محاولات عديدة. يقضي ياسين الآن وقته في المنزل يعمل على الكمبيوتر.

| كثيراً                | 6                     | 5                     | معتدل                 | 4                     | 3                     | 2                     | على الإطلاق           |  |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | ما مدى احصالية إقناع ياسين بالانضمام إلى جماعة ذات معتقدات مختلفة؟ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | ما مدى تكلف في جوابك؟  |

*your answer?*”, see Figure 4. A reliability analysis on multiple outcome variables showed high intercorrelations between the items and thus good internal consistency for the manipulativeness measure ( $\alpha = 0.80$ ,  $M = 3.93$ ,  $SD = 1.08$ ), and for the confidence measure ( $\alpha = 0.89$ ,  $M = 5.08$ ,  $SD = 1.06$ ); see Table S3 and Figure S3 in the SI.

Similarly, after each profile vignette, participants were asked “*How likely is it that [name of the individual] might join a group with different beliefs?*”, and again “*How confident are you in your answer?*” (with response options again ranging from 1 “*not at all*” and 7 “*very*”), see Figure 5. A reliability analysis showed high intercorrelations between the vignettes, indicating good internal consistency ( $\alpha = 0.77$  for the vulnerability measure and  $\alpha = 0.90$  for the confidence measure, see Table S3 and Figure S3 in the SI).

As an exploratory measure, participants were also asked to indicate their level of “*oneness*” (i.e., sense of identification or connectedness) with the following groups on 1-7 Likert scale (1 being “*not at all connected*” and 7 being “*strongly connected*”): the media, family, friends, and neighbours. Finally, participants answered a few demographic questions (age, gender, marital status, and whether they had participated in a Spirit of Soccer workshop before).

### 3. RESULTS

We followed the analysis plan from Saleh et al. (2021). For a full overview of the results, we refer to Table S2 in the SI. With respect to the WhatsApp questions (**H1**), a one-way analysis of covariance (ANCOVA) revealed a statistically significant post-game difference between treatment group ( $M = 4.07$ ,  $SD = 1.12$ ) and control group ( $M = 3.78$ ,  $SD = 1.03$ ) for the perceived manipulativeness of the aggregated index of the WhatsApp messages, controlling for the pre-test scores ( $F(1,188) = 4.55$ ,  $p = 0.034$ ,  $\eta^2 = 0.018$ ). Specifically, the shift in manipulativeness

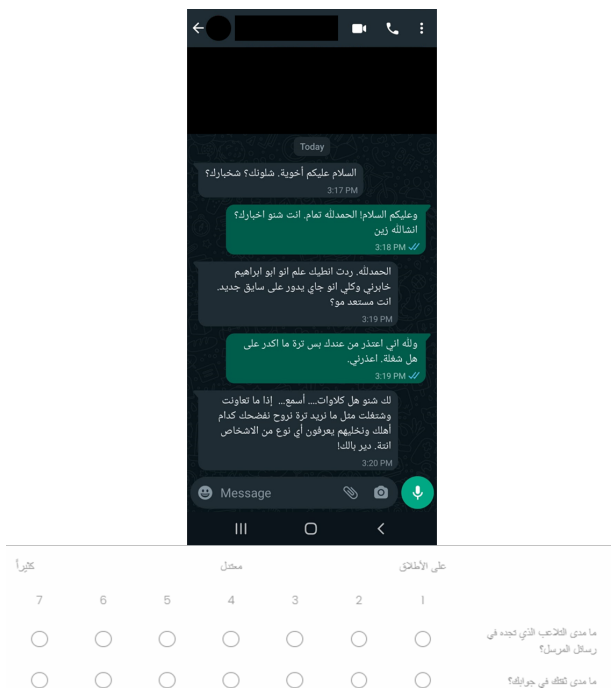
score post-intervention was significantly higher in the inoculation condition than in the control group ( $M_{diff} = -0.29$ , 95% CI [-0.596, -0.021],  $d = 0.31$ ). These results support **H1**, albeit at a smaller effect size than the original study ( $d = 0.71$ ). Figure 6 shows the violin and density plots for the post-gameplay manipulativeness scores for both conditions. See also Figure S1 in the Supplementary Information.

With respect to the confidence measure, a one-way ANCOVA controlling for the pre-test showed a trend where participants in the inoculation group became more confident in their assessment of the manipulativeness of the WhatsApp posts ( $M = 5.24$ ,  $SD = 1.05$ ) compared to the control group ( $M = 4.93$ ,  $SD = 1.05$ ;  $F(1,188) = 3.86$ ,  $p = 0.051$ ,  $\eta^2 = 0.01$ ), with an effect size of  $d = 0.29$ , although this difference did not reach statistical significance. We thus did not find support for **H3**. However, the results fell just above the traditional (0.05) significance level and may be explained by a lack of statistical power rather than the true absence of an effect. Figure 7 shows these results in a series of boxplots. See also Figure S2 in the SI.

For the vignettes measure, we ran a one-way ANCOVA on the aggregated perceived vulnerability of the profile vignettes, with the aggregated pretest score as the covariate. Doing so showed no significant postintervention difference between the inoculation ( $M = 4.00$ ,  $SD = 0.89$ ) and control condition ( $M = 4.00$ ,  $SD = 0.94$ ). Here, we found no effect of the treatment condition on aggregated perceived vulnerability ( $F(1,188) = 0.02$ ,  $p = 0.897$ ,  $\eta^2 = 0.000$ ). This result fails to confirm **H2**.

In Table 2, we compared our results to the original study by Salet et al. (2021). The table shows the key outcome measures, including WhatsApp messages’ manipulativeness and confidence, vignettes’ vulnerability and confidence, and non-vulnerable vignettes’ vulnerability and confidence. One of the main measures, WhatsApp messages’ perceived manipulativeness, is

**Figure 4**  
**Outcome Measures (Perceived Manipulativeness & Confidence) Presented after each WhatsApp Message Stimulus.**



**Figure 5**  
**Outcome Measures (Likelihood of This Person Joining a Group With Different Beliefs & Confidence) Presented After Each Profile Vignette.**



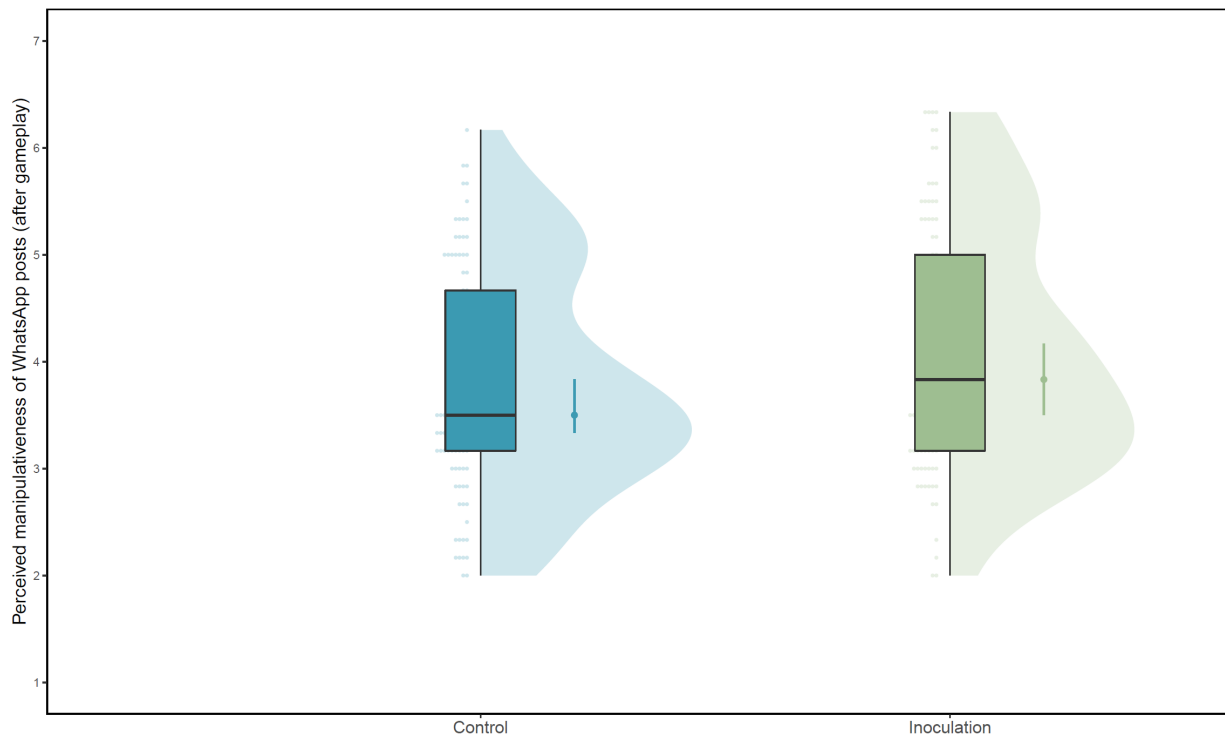
significant in both studies. In addition, the confidence measure for the WhatsApp messages yielded a *p*-value of 0.051, falling just short of being statistically significant. For the vulnerable vignettes, both vulnerability and confidence measures were not statistically significant in the present study compared to the original. Last, we see that for the non-vulnerable vi-

gnettes, both the vulnerability and confidence measures were not statistically significant, as was the case in the original study.

In Table 3, we present the results of a series of (exploratory) Two One-Sided independent-samples *t*-tests (TOSTs), which allows us to explore in more detail if our results represent

**Figure 6**

**Box-violin Plot With Data Jitter and Point-Range Plots for the Perceived Manipulativeness of WhatsApp Messages After Gameplay, for the Control (*Tetris*) and Inoculation (*MindFort*) Conditions.** The distribution is summarised by a boxplot (not showing outliers), a point-range (showing the median and its 95% percentile-bootstrapped confidence interval), density plot, and a dot plot.



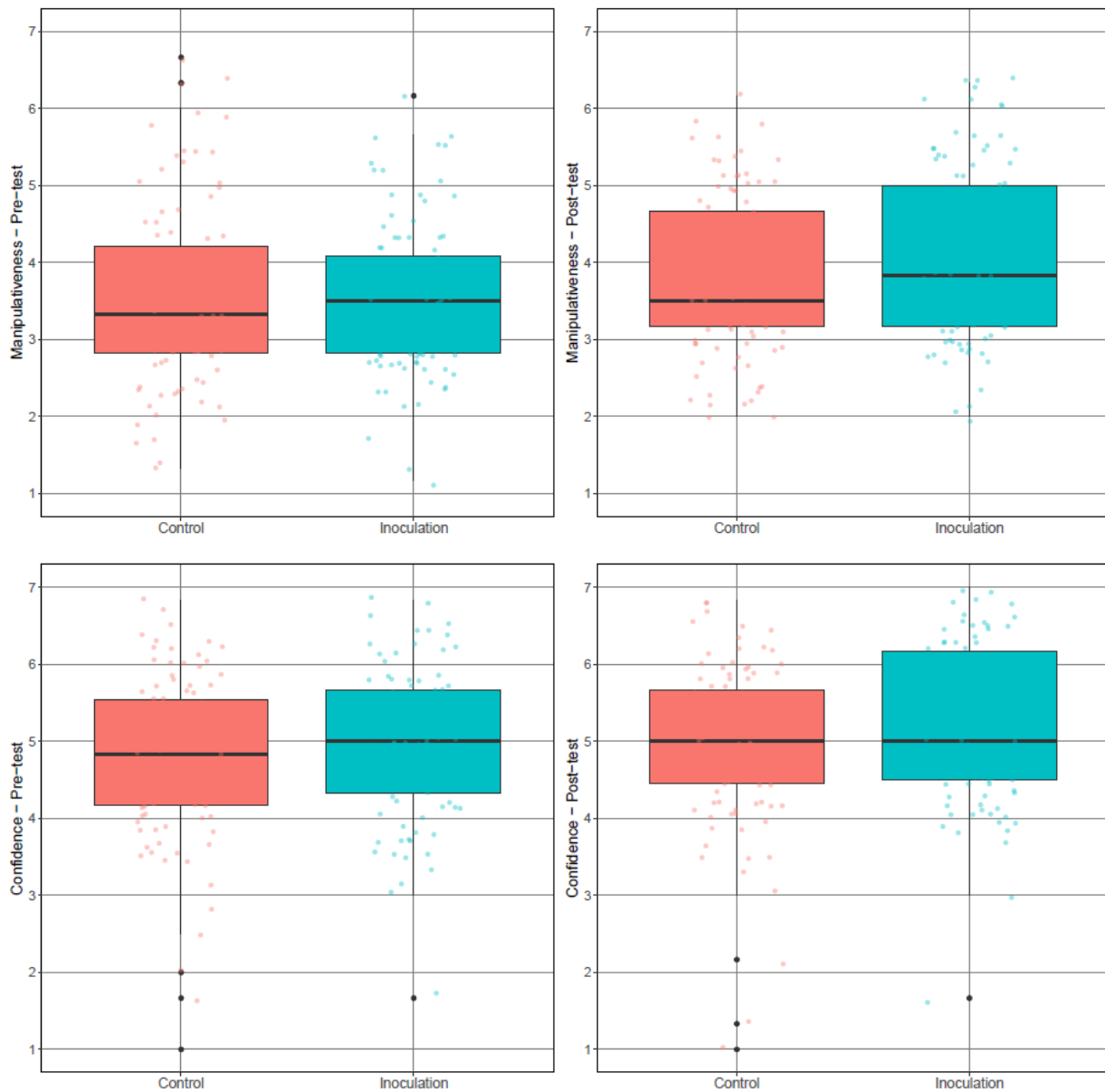
**Table 2**

Comparison of mean values, *p*-values and effect sizes for the present study versus the original study (Saleh et al., 2021).

| Outcome measure                          | Present Study           |          |                  | Original Study          |          |                  |
|--|-------------------------|----------|------------------|-------------------------|----------|------------------|
|  | <i>M<sub>diff</sub></i> | <i>p</i> | Cohen's <i>d</i> | <i>M<sub>diff</sub></i> | <i>p</i> | Cohen's <i>d</i> |
| WhatsApp messages - Manipulativeness     | -0.29                   | 0.034    | 0.31             | -0.57                   | <0.001   | 0.71             |
| WhatsApp messages - Confidence           | 0.22                    | 0.051    | 0.29             | 0.29                    | < 0.001  | 0.41             |
| Vignettes - Vulnerability                | 0.01                    | 0.897    | 0.02             | -0.84                   | <0.001   | 0.81             |
| Vignettes - Confidence                   | -0.11                   | 0.357    | 0.10             | 0.14                    | < 0.001  | 0.45             |
| Non-Vulnerable Vignettes - Vulnerability | 0.12                    | 0.489    | 0.09             | -0.20                   | 0.556    | 0.06             |
| Non-Vulnerable Vignettes - Confidence    | -0.15                   | 0.368    | 0.12             | 0.09                    | 0.235    | 0.11             |

**Figure 7**

**Box Plots for the Perceived Manipulativeness and Confidence in Judgment of WhatsApp Messages Before (pre) and After (post) Gameplay, for the Inoculation (*MindFort*) and Control (*Tetris*) Conditions.**



“true null” effects, or if a potential lack of statistical power means we should not rule out the presence of potentially significant effects at a higher sample size. We chose  $d = 0.30$  as our smallest effect size of interest (SESOI), which means that we explore if we can rule out potentially significant effects at this effect size or larger.

Table 3 shows that the TOST was not significant for the manipulativeness and confidence measures for the WhatsApp messages, and for the vulnerability measure for the non-vulnerable vignettes. This indicates that for these three outcome measures, we cannot rule out the presence of significant effects at Cohen’s  $d = 0.30$  or higher. For the other three measures, both the upper and lower bounds of the TOST are significant, indicating that there

**Table 3**

Two One-Sided Tests (TOSTs; here independent-samples *t*-tests) for the pre-post difference scores for the six outcome variables of interest. Smallest Effect Size of Interest (SESOI) is  $d = 0.30$ .

| Variable                                 |                | <i>t</i> | <i>df</i> | <i>p</i>     |
|--|----------------|----------|-----------|--------------|
| WhatsApp messages - Manipulativeness     | <i>t</i> -test | -1.85    | 187       | 0.066        |
|  | TOST Upper     | 0.03     | 187       | <b>0.489</b> |
|  | TOST Lower     | -3.73    | 187       | < 0.001      |
| WhatsApp messages - Confidence           | <i>t</i> -test | -1.50    | 181       | 0.135        |
|  | TOST Upper     | 1.05     | 181       | <b>0.148</b> |
|  | TOST Lower     | -4.05    | 181       | < 0.001      |
| Vignettes - Vulnerability                | <i>t</i> -test | 0.16     | 188       | 0.877        |
|  | TOST Upper     | 2.74     | 188       | 0.003        |
|  | TOST Lower     | -2.43    | 188       | 0.008        |
| Vignettes - Confidence                   | <i>t</i> -test | -0.38    | 184       | 0.705        |
|  | TOST Upper     | 2.05     | 184       | 0.021        |
|  | TOST Lower     | -2.81    | 184       | 0.003        |
| Non-Vulnerable Vignettes - Vulnerability | <i>t</i> -test | -0.66    | 189       | 0.510        |
|  | TOST Upper     | 0.77     | 189       | <b>0.221</b> |
|  | TOST Lower     | -2.09    | 189       | 0.019        |
| Non-Vulnerable Vignettes - Confidence    | <i>t</i> -test | 0.19     | 189       | 0.853        |
|  | TOST Upper     | 1.77     | 189       | 0.047        |
|  | TOST Lower     | -1.39    | 189       | 0.083        |

Note. Non-significant *p*-values for the TOST upper or lower bound indicates evidence *against* the null hypothesis of no between-group difference at effects larger than the SESOI.

is no expected significant effect at  $d = 0.30$  or higher. This analysis reaffirms our earlier findings: taking into account the lack of power, the inoculation (*MindFort*) group performed better than the control group at assessing the WhatsApp messages, but this was not the case for the vignettes measures.

We also analysed our findings under a Bayesian framework in Table S4 in the SI. We found weak support for **H1** (the WhatsApp messages manipulateness measure;  $BF_{10} = 1.49$ ); all other Bayes factors are lower than 0, indicating varying levels of support for the null hypotheses (namely that the inoculation group mean is not significantly larger than the control group

mean). For the WhatsApp messages confidence measure, the Bayes factor is 0.833, indicating weak support for the null and a need to collect more data. This reaffirms that, particularly for the manipulateness and confidence measures, this study was most likely underpowered to detect significant effects, but the Bayesian analyses did not rule out that strong confidence *against* the null hypothesis may be obtained if more data is collected. However, as discussed, we were unable to do so due to the complications of data collection associated with this particular study design and sample.

With respect to the “Oneness” exploratory measure (i.e., sense of identification/connect-edness with the media, family, friends, and neighbours), although we found that treatment group participants reported a significantly higher sense of Oneness with their friends ( $p = 0.035$ ,  $d = 0.31$ ; see Table S2 in the SI), this result was not significant when adjusting for multiple comparisons. We therefore did not make strong inferences about how the game affects people’s feelings of identification with various groups.

Finally, we conducted a series of linear regressions with the pre-post difference scores for our outcome measures of interest as the dependent variable and experimental condition, age, gender, marital status, and SoS workshop participation as factor variables, to see if our results were robust to the inclusion of covariates (see Table S5 in the SI). We report that the pre-post difference in the perceived manipulateness of WhatsApp messages was significantly higher in the inoculation group than the control group, when controlling for demographics ( $b = 0.33$ ,  $p = 0.035$ ). We found no other significant effects of experimental condition on our outcome measures (all  $p$  values  $> 0.144$ ). In addition, none of the demographic variables were independent significant predictors of the outcome measures, with the exception of age, which was associated with a lower pre-post difference score for the perceived manipulateness of

WhatsApp messages ( $b = -0.03$ ,  $p = 0.044$ ) and a higher pre-post difference in confidence in identifying non-vulnerable vignettes ( $b = 0.05$ ,  $p = 0.005$ ).

#### 4. DISCUSSION

In this conceptual replication of Saleh et al. (2021), we evaluated the effectiveness of a game-based inoculation intervention aimed at reducing psychological susceptibility to manipulation techniques commonly used by extremist organisations in a field study conducted in post-conflict regions of Iraq. We report the following results: despite the difficulties associated with field research in general (DellaVigna & Linos, 2022; Sriram et al., 2009) and doing research in post-conflict zones suffering from low literacy rates (Wood, 2006), we find that playing the *MindFort* game significantly improved people’s ability to identify manipulative messaging. The overall effect size ( $d = 0.31$ ) was significant but substantially smaller than the original study ( $d = 0.71$ ); such a reduction in effect size is to be expected when implementing lab studies in the field (DellaVigna & Linos, 2022; Roozenbeek, Van Der Linden, et al., 2022). In addition, participants’ confidence in their judgments of manipulative content also improved; this effect was not quite significant ( $p = 0.051$ ), but this could be attributed to a lack of sufficient statistical power. Finally, we found no effect of the game on participants’ assessment of the characteristics that make people vulnerable to extremist recruitment, unlike in the original study. Overall, we view these findings (with the exception of the vignettes) as broadly in line with previous research (Saleh et al., 2021).

One of the goals of this study was to address a key limitation in the study by Saleh et al. (2021): not being able to select participants who are vulnerable to extremist recruitment. Compared to the online recruitment of UK-based participants from Prolific Academic, conducting this intervention as a field experiment in a post-conflict region shows the potential impact

of game-based interventions in such areas. A high-level comparison of the mean values of this study to those from Saleh et al. (2021) may help shed light on the observed differences between Iraq and the UK. For the perceived manipulateness (**H1**), the mean values here (treatment group = 4.07, control group = 3.78) were lower than the mean values for the same measures (6.22 and 5.64) reported in Saleh et al. (2021). For the perceived vulnerability measure (**H2**), the present study showed similar mean values (4.00) in both the treatment and control groups, which were lower than the treatment and control groups (5.11 and 4.28, respectively) in the original study. Likewise, the mean values for confidence (**H3**) were lower in both the treatment and control groups (5.24 and 4.93, respectively) in the present study than in the comparable conditions (6.12 and 5.83) in Saleh et al. (2021).

This generally lower level of perceived manipulateness of the WhatsApp messages, and the lack of an effect of the game on participants' ability to identify the factors that make people vulnerable to extremist recruitment (i.e., the vignettes measure, see Table 2), implies that some or all of these stimuli fall short of passing the threshold for individuals from the areas of Iraq where this research was conducted. Compared to the United Kingdom, the WhatsApp messages may thus stand out less from "normal" discourse, which may have reduced responsiveness and blunted perceptions of manipulative messaging. With respect to the vignettes, the complexity of a post-conflict region may make it more difficult for one to pick up on traits that could be seen as "out of the ordinary" in normal circumstances for a country in the Global North. For example, in many Western countries it is often considered a red flag when a student shows less interest in academia, drops out of school, and begins to spend time with a new (and possibly unsavoury) circle of friends. However, in a region where thousands of kids had to stop going to school for sev-

eral years due to displacement, destruction of schools and infrastructure, or ISIS guidelines preventing girls from attending schools, then the above example may not be so out of the ordinary, which could explain the differences between the present conceptual replication and the original study. As mentioned in the Introduction, it is possible that a desensitisation to manipulation attempts in Iraq, possibly due to increased exposure to extremist recruitment attempts or other factors, contributed to this effect (Bushman & Anderson, 2009; Mrug et al., 2016). This may suggest that participants had higher "prior exposure" levels than anticipated, which may have reduced the efficacy of the inoculation.

It is also possible that the adaptation and translation of the game and stimuli affected participants' responses. Because the experiment was tested in volatile and sensitive areas in Iraq, it was important to sanitise the content and translate it in a way that avoids mentioning certain words that could be misperceived by the target population (i.e., words and phrases related to radicalisation, terrorism, or violence). This may have made it more difficult for participants to draw a connection between certain traits mentioned in the vignettes and WhatsApp messages, and the risk of being recruited into an extremist organisation.

A final explanation for the weaker effects compared to the original study lies in the fact that the control group also showed a non-zero pre-post improvement. Treatment group participants' assessment of the manipulateness of the WhatsApp messages ( $M_{post} = 4.07$ ,  $SD_{post} = 1.12$ ) increased by around 15% compared to pre-game assessment ( $M_{pre} = 3.53$ ,  $SD_{pre} = 0.99$ ). However, the control group also rated WhatsApp messages as more manipulative, albeit at about 7% ( $M_{pre} = 3.53$ ,  $SD_{pre} = 1.18$ ;  $M_{post} = 3.78$ ,  $SD_{post} = 1.03$ ); see Figure 7 and Figure S1 in the SI. This may be attributed to various factors related to field testing, including noise in the data and/or contamination happening



during the implementation, since on some occasions control group participants carried out the experiment within the same space as the treatment group.

It is important to note that the bulk of radicalisation and extremism research has been conducted in the context of countries in the Global North. However, the psychological factors that underlie vulnerability to extremism may not be universally applicable across different regions and cultures (Ghai, 2021). Literature from Western countries identifies factors such as unemployment (Altunbas & Thornton, 2011; Bouhana & Wikström, 2011; Gartenstein-Ross & Grossman, 2009), incarceration (Brandon, 2009; Rushchenko, 2018), disconnection from society (Bartlett et al., 2010; Brandon, 2009; De Poot & Sonnenschein, 2011; Leiken & Brooke, 2006), identity seeking (Dalgaard-Nielsen, 2010), and adventure seeking (Borárosová et al., 2017) as possible drivers for radicalisation. In contrast, there is considerable research showing that factors such as threat, relative deprivation, perceived injustice, and identity fusion may fuel extremism in the Global South (Charkawi et al., 2021; Obaidi et al., 2019; Van Den Bos, 2020). By including participants from hard-to-reach areas and studying how vulnerability to extremist recruitment may be reduced, this study has offered an entry point for a more comprehensive understanding of extremism and how to counter it outside of Western countries. However, further research is needed to understand the specific nuances in terms of characteristics that contribute to extremist recruitment, taking into account the unique socio-political and cultural contexts of different regions.

There are several potential pathways forward for research into the dynamics of extremism and interventions to counter the rhetorical strategies used in extremist recruitment. First, although not always feasible, it is important that studies conducted in hard-to-reach areas are high-powered. As this may impose substantial challenges on research and data collection

teams, it may be worth considering to what extent pooled funding or other avenues of collaboration are an option. Second, sample diversity and its consequences not only for the comparability of studies across countries and contexts, but also with respect to study design, must be actively considered (Ghai, 2021, 2022). Third, researchers may consider testing *why* exactly interventions might yield different outcomes across countries, contexts, and samples, for instance by conducting the same study in multiple countries and including a larger battery of scales that may shed light on the underlying mechanisms of intervention effectiveness (such as perceived motivation and objective memory etc). Within the context of violent extremism, we further suggest exploring whether factors such as poverty, unemployment, disconnection from society, trust in the government and other relevant institutions, and desensitisation play similar or different roles depending on the context.

## 5. CONCLUSION

While it is often challenging to replicate results from lab-based experiments in the field, results from this study conducted in post-conflict areas of Iraq show that playing a short game about radicalisation can improve people's ability to recognise manipulative messaging and commonly used strategies to recruit people into extremist organisations. Such tools may be applied at scale, for example as part of educational programmes or workshops, and add much-needed evidence for "what works" in the context of preventing violent extremism. However, we note that this study was most likely underpowered. More research is therefore needed to address the limitations imposed by conducting research in difficult-to-access areas.

## 6. DECLARATIONS OF INTEREST

We have no known conflict of interest to disclose.



## 7. ACKNOWLEDGEMENTS

Special thanks go to our implementation partners, Spirit of Soccer, for dedicating resources, offering extensive help, and navigating us through the volatilities of some of the most perilous regions of Iraq. They stand true to their vision: educating, equipping, and employing vulnerable people in conflict and post-conflict zones around the world. We would also like to thank Google for their support and specifically Tanner Verhey for his excellent feedback.

## FOOTNOTES

- <sup>1</sup> We did not preregister this study but instead follow the hypotheses and analysis plan from the original study, which was preregistered.

## REFERENCES

- Altunbas, Y., & Thornton, J. (2011). Are home-grown Islamic terrorists different? Some UK evidence. *Southern Economic Journal*, 78(2), 262–272. <https://doi.org/10.4284/0038-4038-78.2.262>
- Amirali, A. (2019). The “Youth Bulge” and Political Unrest in Iraq: A Political Economy Approach. *UK: Institute of Development Studies*.
- Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on Inoculation Theory. *Communication Monographs*, 77(3), 281–311. <https://doi.org/10.1080/03637751003758193>
- Bartlett, J., Birdwell, J., & King, M. (2010). *The edge of violence: A radical approach to extremism*. Demos.
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W., & Van Der Linden, S. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data and Society*, 8(1). <https://doi.org/10.1177/20539517211013868>
- Basol, M., Roozenbeek, J., & Van Der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, 3(1), 1–9. <https://doi.org/10.5334/joc.91>
- BBC News. (2016). *Islamic State “loses 40% of territory in Iraq*. <https://www.bbc.com/news/world-middle-east-35231664> Retrieved December 12, 2022.
- Berger, J. M., & Morgan, J. (2015). The ISIS Twitter census. Defining and describing the population of ISIS supporters on Twitter. *The Brookings Project on U.S. Relations with the Islamic World*. <https://www.brookings.edu/articles/the-isis-twitter-census-defining-and-describing-the-population-of-isis-supporters-on-twitter/> (No. 20).
- Borárosová, I., Walter, A., & Filipec, O. (2017). *Global Jihad: Case studies in terrorist organizations*. Research Institute for European Policy.
- Bouhana, N., & Wikström, P.-O. H. (2011). *Al Qa’ida-influenced radicalisation: A rapid evidence assessment guided by Situational Action Theory*. Home Office.
- Brandon, J. (2009). The danger of prison radicalization in the West. *CTC Sentinel*, 2(12). <https://ctc.westpoint.edu/the-danger-of-prison-radicalization-in-the-west/>
- Bushman, B. J., & Anderson, C. A. (2009). Comfortably numb: Desensitizing effects of violent media on helping others. *Psychological Science*, 20(3), 273–277. <https://doi.org/10.1111/j.1467-9280.2009.02287.x>
- Charkawi, W., Dunn, K., & Bliuc, A. M. (2021). The influences of social identity and perceptions of injustice on support to violent extremism. *Behavioral Sciences of Terrorism and Political Aggression*, 13(3), 177–196. <https://doi.org/10.1080/19434472.2020.1734046>
- Compton, J. (2021). Threat and/in Inoculation Theory. *International Journal of Communication*, 15, 4294–4306. <https://ijoc.org/index.php/ijoc/article/view/17634/3565>
- Compton, J., Jackson, B., & Dimmock, J. A. (2016). Persuading others to avoid persuasion: Inoculation Theory and resistant health attitudes. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00122>



- Compton, J., Van Der Linden, S., Cook, J., & Basol, M. (2021). Inoculation Theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass*, 15(6), e12602. <https://doi.org/10.1111/spc3.12602>
- Cook, J., Ecker, U. K. H., Trecek-King, M., Schade, G., Jeffers-Tracy, K., Fessmann, J., Kim, S. C., Kinkead, D., Orr, M., Vraga, E., Roberts, K., & McDowell, J. (2022). The Cranky Uncle game—Combining humor and gamification to build student resilience against climate misinformation. *Environmental Education Research*. <https://doi.org/10.1080/13504622.2022.2085671>
- Dalgaard-Nielsen, A. (2010). Violent radicalization in Europe: What we know and what we do not know. *Studies in Conflict & Terrorism*, 33(9), 797–814. <https://doi.org/10.1080/1057610X.2010.501423>
- De Poot, C. J., & Sonnenschein, A. (2011). *Jihadi terrorism in the Netherlands: A description based on closed criminal investigations*. Boom Juridische Uitgevers.
- DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1), 81–116. <https://doi.org/10.3982/ECTA18709>
- Doosje, B., Moghaddam, F. M., Kruglanski, A. W., De Wolf, A., Mann, L., & Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11, 79–84. <https://doi.org/10.1016/j.copsyc.2016.06.008>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich.
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Euer, K., Van Vossle, A., Groenen, A., & Van Bouchaute, K. (2014). *Strengthening resilience against violent radicalization (STRE-SAVIORA) Part I: Literature analysis*. Federal Public Service Home Affairs.
- Gartenstein-Ross, D., & Grossman, L. (2009). *Homegrown terrorists in the U.S. and the U.K.: An empirical examination of the radicalization process*. FDD Press.
- Ghai, S. (2021). It's time to reimagine sample diversity and retire the WEIRD dichotomy. *Nature Human Behaviour*, 5(8), 971–972. <https://doi.org/10.1038/s41562-021-01175-9>
- Ghai, S. (2022). Expand diversity definitions beyond their Western perspective. *Nature*, 602(7896), 211–211. <https://doi.org/10.1038/d41586-022-00330-0>
- Glenn, C., Mattisan, R., Caves, J., & Nada, G. (2019). *Timeline: The rise, spread, and fall of the Islamic State*. <https://www.wilsoncenter.org/article/timeline-the-rise-spread-and-fall-the-islamic-state>
- Green, M., Mcshane, C. J., & Swinbourne, A. (2022). Active versus passive: evaluating the effectiveness of inoculation techniques in relation to misinformation about climate change. *Australian Journal of Psychology*, 74(1). <https://doi.org/10.1080/00049530.2022.2113340>
- Hogg, M. A., Kruglanski, A. W., Van Den, & Bos, K. (2013). Uncertainty and the roots of extremism. *Journal of Social Issues*, 69(3), 407–418. <https://doi.org/10.1111/josi.12021>
- Institute for Economics & Peace. (2022). *Global Terrorism Index - Measuring the impact of terrorism*. <https://www.visionofhumanity.org/wp-content/uploads/2022/03/GTI-2022-web-04112022.pdf> Retrieved December 12, 2022.
- Ivanov, B., Parker, K. A., & Dillingham, L. L. (2018). Testing the limits of inoculation-generated resistance. *Western Journal of Communication*, 82(5), 648–665. <https://doi.org/10.1080/10570314.2018.1454600>
- Kemp, S. (2022). *Digital 2022: Iraq*. [https://datareportal.com/reports/digital-2022-iraq#:~:text=Internet%20use%20in%20Iraq%20in,percent\)%20between%202021%20and%](https://datareportal.com/reports/digital-2022-iraq#:~:text=Internet%20use%20in%20Iraq%20in,percent)%20between%202021%20and%)

- 202022 Retrieved November 18, 2022.
- Kittleson, S. (2020). *Islamic State conducts attacks near Iraq's Syrian and Iranian borders*. <https://www.al-monitor.com/originals/2020/04/iraq-isis-diyala-jelawla-terrorism-security.html> Retrieved December 12, 2022.
  - Knudsen, R. A. (2018). Measuring radicalisation: Risk assessment conceptualisations and practice in England and Wales. *Behavioral Sciences of Terrorism and Political Aggression*, 1–18. <https://doi.org/10.1080/19434472.2018.1509105>
  - Kouachi, I. I. (2022). *4 policemen killed in Daesh/ISIS attack in Iraq*. <https://www.aa.com.tr/en/middle-east/4-policemen-killed-in-daesh-isis-attack-in-iraq/2641418> Retrieved December 12, 2022.
  - Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology*, 35(S1), 69–93. <https://doi.org/10.1111/pops.12163>
  - Lees, J., Banas, J. A., Linvill, D., Meirick, P. C., & Warren, P. (2023). The Spot the Troll Quiz game increases accuracy in discerning between real and inauthentic social media accounts. *PNAS Nexus*, 2(4), pgad094. <https://doi.org/10.1093/pnasnexus/pgad094>
  - Leiken, R. S., & Brooke, S. (2006). The quantitative analysis of terrorism and immigration: An initial exploration. *Terrorism and Political Violence*, 18(4), 503–521. <https://doi.org/10.1080/09546550600880294>
  - Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
  - Linden, S. V. D., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global challenges*, 1(2), 1600008–1600008.
  - Lu, C., Hu, B., Li, Q., Bi, C., & Ju, X. D. (2023). Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 25, 49255–49255. <https://doi.org/10.2196/49255>
  - Maertens, R., Roozenbeek, J., Basol, M., & Van Der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1–16. <https://doi.org/10.1037/xap0000315>
  - Maertens, R., Roozenbeek, J., Simons, J., Lewandowsky, S., Maturo, V., Goldberg, B., & Van Der Linden, S. (2023). Psychological booster shots targeting memory increase long-term resistance against misinformation. *PsyArxiv Preprints*. <https://psyarxiv.com/6r9as/>
  - McGuire, W. J. (1964). Some contemporary approaches. *Advances in Experimental Social Psychology*, 1(C), 191–229. [https://doi.org/10.1016/S0065-2601\(08\)60052-0](https://doi.org/10.1016/S0065-2601(08)60052-0)
  - McGuire, W. J., & Papageorgis, D. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology*, 63, 326–332. <https://doi.org/10.1037/h0048344>
  - McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly*, 26(1), 24–34. <https://doi.org/10.1086/267068>
  - Mrug, S., Madan, A., & Windle, M. (2016). Emotional desensitization to violence contributes to adolescents' violent behavior. *Journal of Abnormal Child Psychology*, 44(1), 75–86. <https://doi.org/10.1007/s10802-015-9986-x>
  - Obaidi, M., Bergh, R., Akrami, N., & Anjum, G. (2019). Group-based relative deprivation explains endorsement of extremism among western-born Muslims. *Psychological Science*, 30(4), 596–605. <https://doi.org/10.1177/>

- 0956797619834879
- O'mahony, C., Brassil, M., Murphy, G., & Linehan, C. (2023). The efficacy of interventions in reducing belief in conspiracy theories: A systematic review. *Plos One*, 18(4), e0280902. <https://doi.org/10.1371/journal.pone.0280902>
  - Ozer, S., & Bertelsen, P. (2018). Capturing violent radicalization: Developing and validating scales measuring central aspects of radicalization. *Scandinavian Journal of Psychology*, 59(6), 653–660. <https://doi.org/10.1111/sjop.12484>
  - Precht, T. (2007). *Homegrown terrorism and Islamist radicalisation in Europe*. Copenhagen. [https://www.justitsministeriet.dk/sites/default/files/media/Arbejdsomraader/Forskning/Forskningspuljen/2011/2007/Home\\_grown\\_terrorism\\_and\\_Islamist\\_radicalisation\\_in\\_Europe\\_-\\_an\\_assessment\\_of\\_influencing\\_factors\\_2\\_.pdf](https://www.justitsministeriet.dk/sites/default/files/media/Arbejdsomraader/Forskning/Forskningspuljen/2011/2007/Home_grown_terrorism_and_Islamist_radicalisation_in_Europe_-_an_assessment_of_influencing_factors_2_.pdf)
  - Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering misinformation: Evidence, knowledge gaps, and implications of current interventions. *European Psychologist*. <https://doi.org/10.31234/osf.io/b52um>
  - Roozenbeek, J., Traberg, C. S., & Van Der Linden, S. (2022). Technique-based inoculation against real-world misinformation. *Royal Society Open Science*, 9(211719). <https://doi.org/10.1098/rsos.211719>
  - Roozenbeek, J., & Van Der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Humanities and Social Sciences Communications*, 5(65), 1–10. <https://doi.org/10.1057/s41599-019-0279-9>
  - Roozenbeek, J., & Van Der Linden, S. (2020). Breaking Harmony Square: A game that “inoculates” against political misinformation. *The Harvard Kennedy School (HKS) Misinformation Review*, 1(8). <https://doi.org/10.37016/mr-2020-47>
  - Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34). <https://doi.org/10.1126/sciadv.abo6254>
  - Rushchenko, J. (2018). *Prison management of terrorism-related offenders: is separation effective?* <https://henryjacksonsociety.org/publications/prison-management-of-terrorism-related-offenders-is-separation-effective/> Centre for the response to radicalisation and terrorism.
  - Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, 83(3), 526–541. <https://doi.org/10.1037/0022-3514.83.3.526>
  - Saleh, N., et al. (2023). *OSF Repository*. [https://osf.io/89mz6/?view\\_only=1523a95805b0467c947a57fe5f7a6ac0](https://osf.io/89mz6/?view_only=1523a95805b0467c947a57fe5f7a6ac0)
  - Saleh, N., Roozenbeek, J., Makki, F., McClanahan, W., & Van Der Linden, S. (2021). Active inoculation boosts attitudinal resistance against extremist persuasion techniques - A novel approach towards the prevention of violent extremism. *Behavioural Public Policy*, 1–24. <https://doi.org/10.1017/bpp.2020.60>
  - Sriram, C. L., Martin-Ortega, O., King, J. C., Mertus, J. A., & Herman, J. (2009). *Surviving Field Research*. Routledge. <https://doi.org/10.4324/9780203875278>
  - Starr, B., & Tawfeeq, M. (2020). *ISIS carried out attacks on Iraq-Syria border posts after US suspended operations*. <https://edition.cnn.com/2020/01/14/politics/isis-attacks-iraq-syria-border/index.html> Retrieved December 12, 2022.
  - Stein, A. (2017). *Terror, love and brainwashing: Attachment in cults and totalitarian systems*. Routledge.
  - Swann, W. B., Jetten, J., Gomez, A., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, 119(3), 441–456. <https://doi.org/10.1037/a0028589>
  - Thornton, A., & Bouhana, N. (2017). Pre-

- venting radicalization in the UK: Expanding the knowledge-base on the channel programme. *Policing*, 13(3), 331–344. <https://doi.org/10.1093/police/pax036>
- Traberg, C. S., Roozenbeek, J., & Van Der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 136–151. <https://doi.org/10.1177/00027162221087936>
  - UNODC. (2018). *Drivers of violent extremism*. <https://www.unodc.org/e4j/en/terrorism/module-2/key-issues/drivers-of-violent-extremism.html> Retrieved November 15, 2022.
  - Van Den Bos, K. (2020). Unfairness and radicalization. *Annual Review of Psychology*, 71(1), 563–588. <https://doi.org/10.1146/annurev-psych-010419-050953>
  - Van Der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifeneder, M. Jaffé, E. Newman, & N. Schwarz (Eds.), *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation* (pp. 147–170). Routledge. <https://doi.org/10.4324/9780429295379-11>
  - Walters, T. K., Monaghan, R., & Ramírez, J. M. (2013). *Radicalization, terrorism and conflict*. Newcastle: Cambridge Scholars Publishing.
  - Wood, E. J. (2006). The ethical challenges of field research in conflict zones. *Qualitative Sociology*, 29, 373–386. <https://doi.org/10.1007/s11133-006-9027-8>
  - World Bank. (2020). *Iraq: Engaging youth to rebuild the social fabric in Baghdad*. <https://www.worldbank.org/en/news/feature/2020/12/02/iraq-engaging-youth-to-rebuild-the-social-fabric-in-baghdad> Retrieved November 18, 2022.