Credibility as a Double-Edged Sword: How Attributing News to a Trusted Source Affects Users' Responses to Misinformation on Personal Messaging
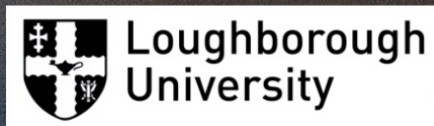
Cristian Vaccari (University of Edinburgh)
Andrew Chadwick (Loughborough University)
Natalie-Anne Hall (Cardiff University)
Brendan Lawson (Loughborough University)

APSA 2024 Annual Meeting

Millions depend on private messaging apps to keep in touch. They're ripe with misinformation

Harris Campaign Starts a WhatsApp Channel to Target Latino Voters

More than half of U.S. Latinos are estimated to regularly use the app to send texts and make calls, but it has also been a major source of misinformation.

Full Fact launches a WhatsApp fact checking service in the UK

Personal Messaging and Misinformation

WhatsApp has >2 billion users globally.

In the UK, 79% of internet users are WhatsApp users.

In many countries, personal messaging apps are more frequently used than public social media platforms.

Online personal messaging is a hard and distinctive misinformation problem due to its social uses and technological affordances.

Deceptive Source Misattribution

The sharing of content that appears to come from a reputable source but is fabricated, altered, or taken out of context.

Takes less effort than other disinformation strategies, such as building an authentic-looking false news website.

Both organized actors and lone individuals are known to use this misleading tactic.

The fake Telegraph article that was shared by Elon Musk – and went viral

A doctored piece claiming rioters would be sent to detention camps was shared on X, formerly Twitter, where it was seen by millions

Ed Cumming
8 August 2024 • 6:00pm

Elon Musk ✔ ✕ @elonmusk · 15m
"Detainment Camps" ...

Ashlea Simon ✔ @AshleaSimonBF · 8h
We're all being deported to the Falklands 😅😂

Keir Starmer considering building 'emergency detainment camps' on the Falkland Islands

The camps would be used to detain prisoners from the ongoing riots as the British prison system is already at capacity

Fiona Parker, SENIOR NEWS REPORTER
7 August 2024 • 11:21am

💬 1.7K   🔁 1.9K   ♡ 9.2K   �archived 974K   🔖 ⬆

Musk, amongst others, shared the false story

At the most superficial level, glimpsed in a flash on social media, it looked like a legitimate *Telegraph* piece.

"Keir Starmer considering building 'emergency detainment camps' on the Falkland Islands," said the headline.

Beneath it, the subheading read: "The camps would be used to detain prisoners from the ongoing riots as the British prison system is already at capacity."

The byline attributed the story to a real *Telegraph* employee. The fonts and formatting were accurate. There was even a plausible date stamp.

But the story was utterly false. Keir Starmer has considered no such thing in response to the unrest that has swept Britain's streets over the past week and a half. Nor has anyone at *The Telegraph* written anything of the kind.

Why Deceptive Source Misattribution Can Be a Problem on Personal Messaging

Personal Messaging: Hybrid public-interpersonal environments, combining personal and public communication between individuals and groups of different sizes.

Misinformation on Personal Messaging: Automated techniques to flag or remove false content are not feasible, making user behavior crucial.

Source Credibility: People rely on source credibility as a shortcut; established news organizations (such as the BBC in the UK) are seen as more credible.

Truth Discernment: Most users can discern between true and false news, but deceptive source misattribution may weaken discernment.

Relevance: Many recent examples show malicious actors rely on deceptive source misattribution to spread falsehoods.

Hypotheses

Independent Variables
Presence of an attributed reputable news source in a message (H1)
Factual accuracy of the news reported in a message (H2)
Interaction between presence of a news source and factual accuracy of the message (H3)

Dependent Variables
Perceived message accuracy
Likelihood of sharing or forwarding the message
Likelihood of staying silent
Likelihood of asking for additional information
Likelihood of providing additional information
Likelihood of verifying the accuracy of the message

Research Design

Pre-registered between-subjects full factorial experiment.

Online survey of 2,580 UK WhatsApp users recruited to match key characteristics of the UK adult population (participation rate 33%).

Participants saw brief vignettes featuring realistic mock-ups of WhatsApp messages with random assignment to messages containing

- True OR false news
- Attributed to BBC News OR no source
- On climate change OR voter ID laws (both topical at the time)
- (Other manipulations that yielded null results: tie strength & group size.)

Dependent variables measured post-exposure.

Debriefing with corrections of misinformation by reputable fact-checkers.
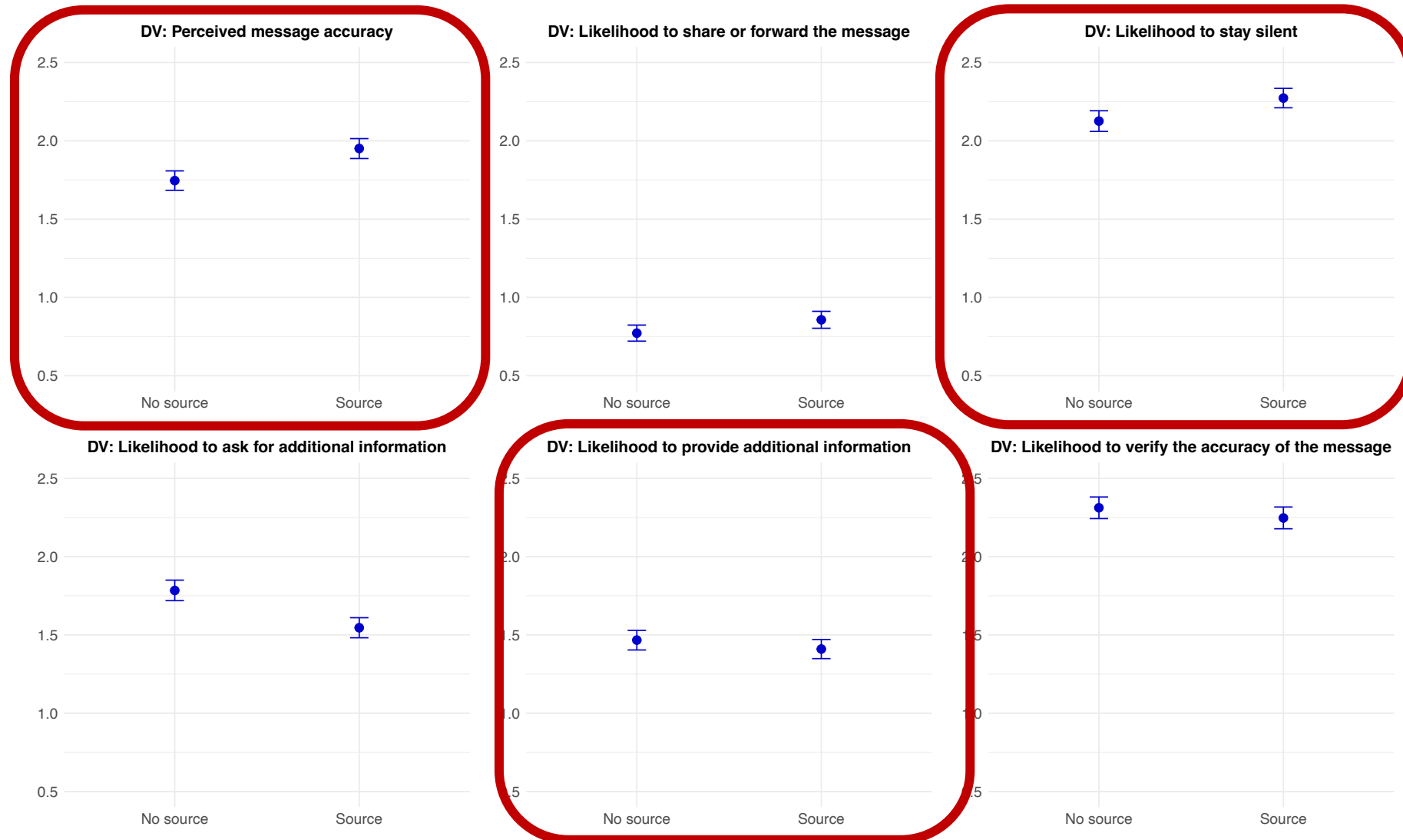
Data analysis: One-way (H1-2) and two-way (H3) ANOVA w/Bonferroni corrections.

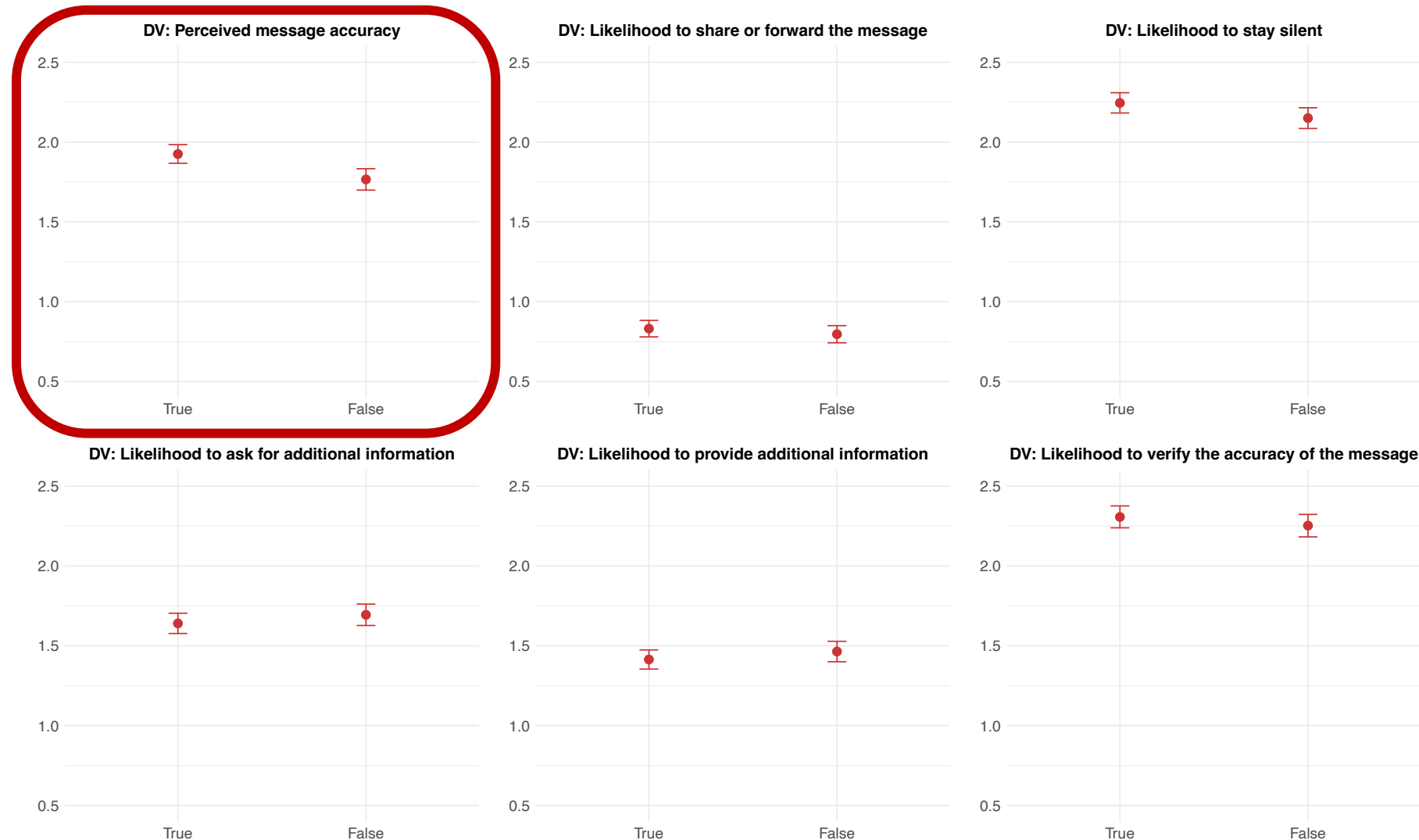# Examples of WhatsApp Mock-Ups in Vignettes

| True Without source | True With source | False Without source | False With source |
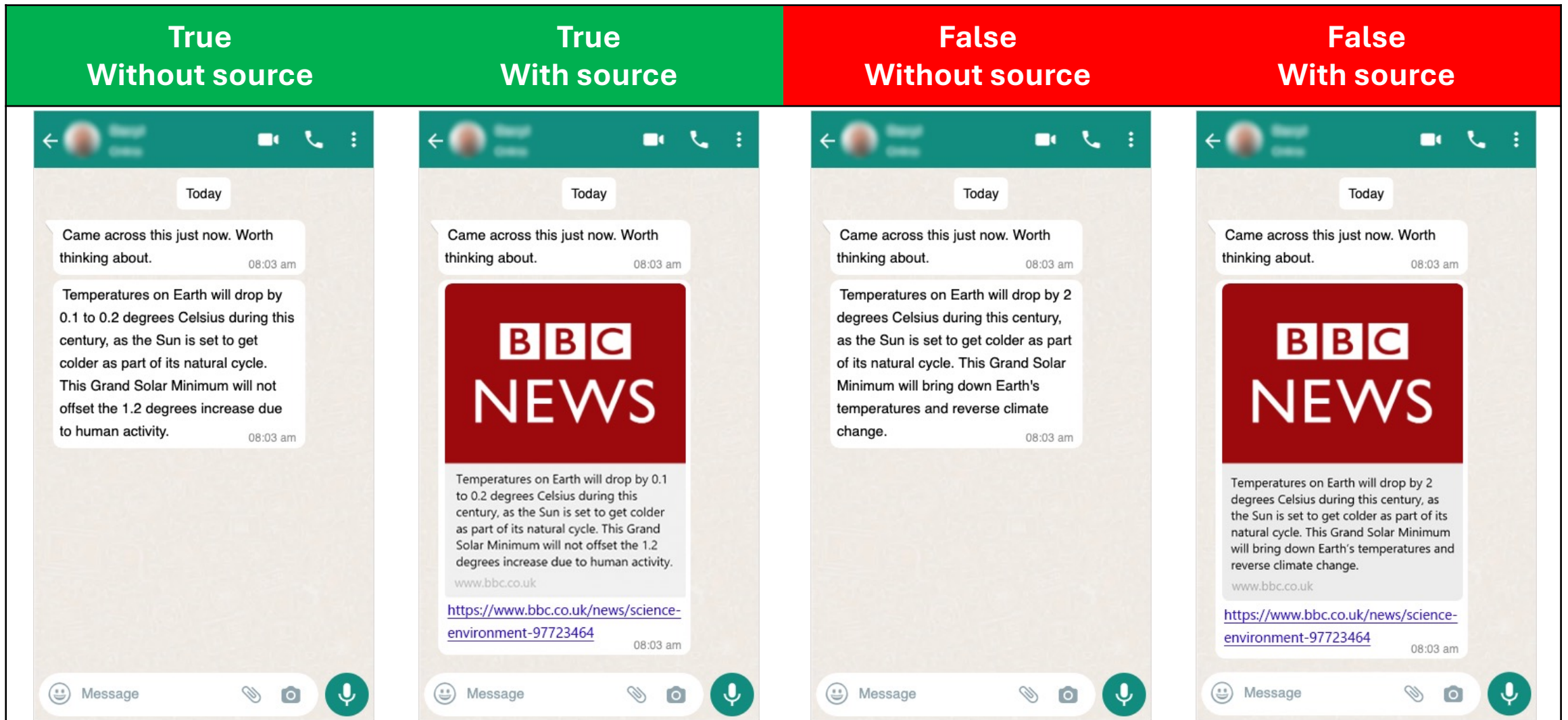
Effects of Seeing a Message with an Attributed Reputable News Source vs No Source (H1)

DV: Perceived message accuracy

DV: Likelihood to share or forward the message

DV: Likelihood to stay silent

DV: Likelihood to ask for additional information

DV: Likelihood to provide additional information

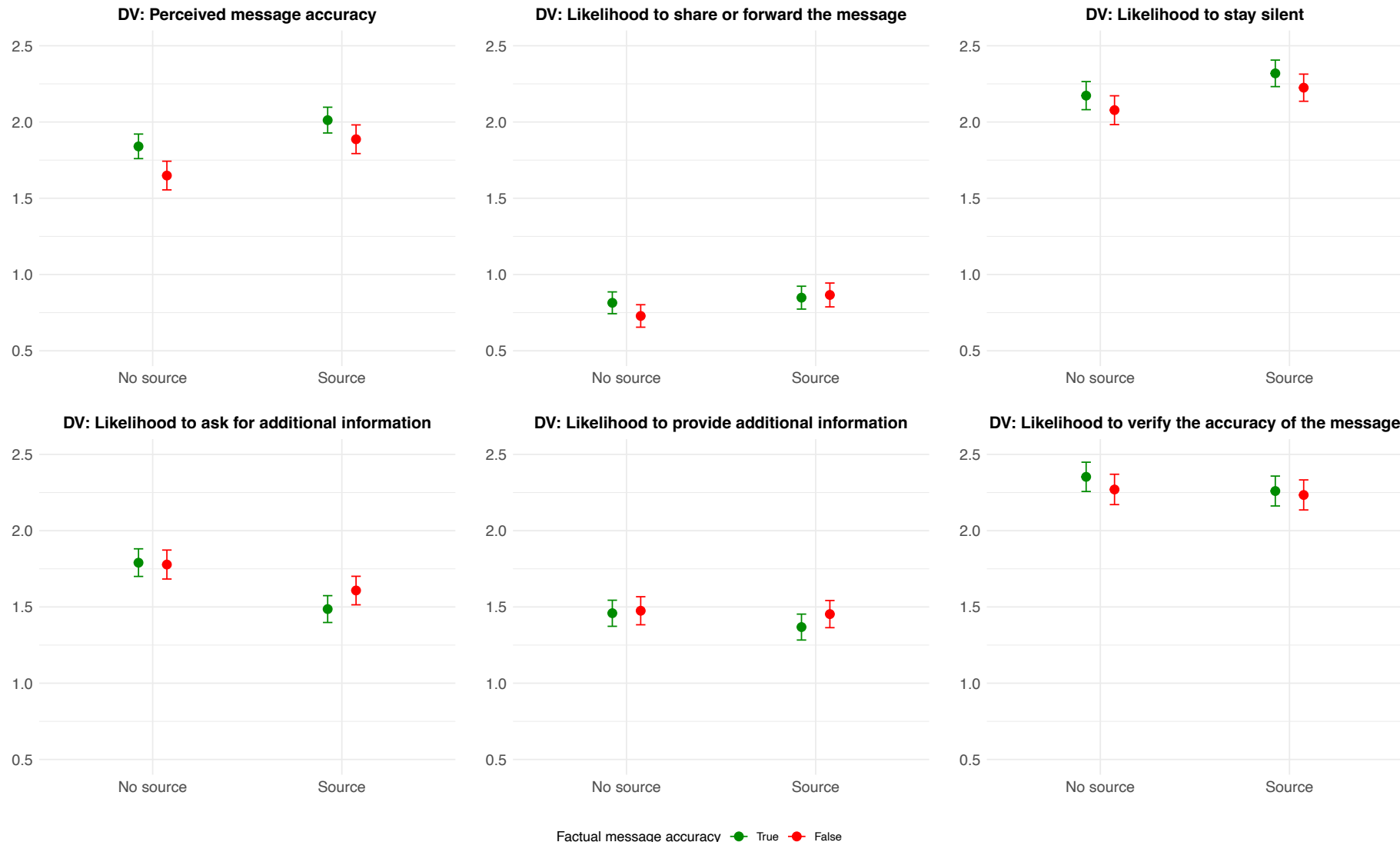**DV: Likelihood to verify the accuracy of the message**

Seeing the BBC News source attributed to information in the WhatsApp message had a statistically significant effect on whether the message was **perceived as accurate**, the likelihood that a participant would **stay silent,** and the likelihood that a participant would **ask for additional information**.

Effects of Seeing a Message Containing News that is Factually Accurate vs Inaccurate (H2)

DV: Perceived message accuracy

DV: Likelihood to share or forward the message

DV: Likelihood to stay silent

DV: Likelihood to ask for additional information

DV: Likelihood to provide additional information

DV: Likelihood to verify the accuracy of the message

Participants randomly assigned to true messages were significantly more likely to **perceive the message as accurate**. They also reported intended behaviors consistent with these perceptions, but none of the effects were statistically significant.

Interaction Between News Source Attribution and Factual Accuracy of the News (H3)

DV: Perceived message accuracy

DV: Likelihood to share or forward the message

DV: Likelihood to stay silent

DV: Likelihood to ask for additional information

DV: Likelihood to provide additional information

DV: Likelihood to verify the accuracy of the message

Factual message accuracy ● True ● False

Attributing news to the BBC in the WhatsApp message produced the same effects for both true and false messages.

When they saw false news attributed to a reputable source, participants reacted in essentially the same way as when they saw true news.

Key Takeaways

1. Attribution to a reputable source elevated the credibility of **true and false news alike**—and thus failed to enhance participants' discernment between true and false information.

Deceptive Source Misattribution on personal messaging can work.

2. The effects of source attribution include both **beliefs** in the veracity of the information and **behaviors** that affect its spread.

3. Personal messaging users can **discern** between true and false news but may not be prepared to **act** upon their judgments.

Limitations

- Results may not generalize to other platforms beyond WhatsApp.
- Simulation of intended behavior in a controlled experiment.
- Focus on the UK and a single news source.
- Self-reports of participants' intended behaviors.
- Small effects sizes.

Key Implications

For Digital Platforms: Need for policies against deceptive source misattribution.

For Policymakers: Support Public Service Media to protect their reputation online.

For News Organizations: Protecting brand integrity and credibility is essential, especially in private messaging contexts.

For Researchers: Investigate how the credibility of reputable sources, while spreading accurate information, can also be exploited to enhance disinformation in personal messaging.

APSA 2024
Online

7 September 2024

Cristian Vaccari
THANK YOU!!! 😀 05:08 pm

Cristian Vaccari
Questions ❓ 05:08 pm

Cristian Vaccari
📬 cvaccari@ed.ac.uk 05:08 pm

Cristian Vaccari
🎙️ https://cristianvaccari.com/
05:08 pm

The Everyday Misinformation Project
🔍 https://everyday-mis.info/ 05:08 pm

Message