

WestminsterResearch

<http://www.westminster.ac.uk/westminsterresearch>

Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation

Buchanan, T.

This is a copy of the accepted author manuscript of an article to be published in PLoS ONE.

The final published version will be available from the publisher at:

<https://journals.plos.org/plosone/>

Copyright © 2020 Prof Tom Buchanan This is an open access article distributed under the Creative Commons Attribution License (CC-BY) (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

1
2
3
4
5
6
7
8
9
10
11
12

Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation.

Tom Buchanan^{1*}

¹School of Social Sciences, University of Westminster, London, U.K.

*Corresponding author

E-mail: T.Buchanan@westminster.ac.uk (TB)

1 **Abstract**

2 Individuals who encounter false information on social media may actively spread it
3 further, by sharing or otherwise engaging with it. Much of the spread of
4 disinformation can thus be attributed to human action. Four studies (total $N=2,634$)
5 explored the effect of message attributes (authoritativeness of source, consensus
6 indicators), viewer characteristics (digital literacy, personality, and demographic
7 variables) and their interaction (consistency between message and recipient beliefs)
8 on self-reported likelihood of spreading examples of disinformation. Participants also
9 reported whether they had shared real-world disinformation in the past. Reported
10 likelihood of sharing was not influenced by authoritativeness of the source of the
11 material, nor indicators of how many other people had previously engaged with it.
12 Participants' level of digital literacy had little effect on their responses. The people
13 reporting the greatest likelihood of sharing disinformation were those who thought it
14 likely to be true, or who had pre-existing attitudes consistent with it. They were likely
15 to have previous familiarity with the materials. Across the four studies, personality
16 (lower Agreeableness and Conscientiousness, higher Extraversion and Neuroticism)
17 and demographic variables (male gender, lower age and lower education) were
18 weakly and inconsistently associated with self-reported likelihood of sharing. These
19 findings have implications for strategies more or less likely to work in countering
20 disinformation in social media.

1 Introduction

2 Disinformation is currently a critically important problem in social media and
3 beyond. Typically defined as “the deliberate creation and sharing of false and/or
4 manipulated information that is intended to deceive and mislead audiences, either for
5 the purposes of causing harm, or for political, personal or financial gain”, political
6 disinformation has been characterized as a significant threat to democracy [1, p.10].
7 It forms part of a wider landscape of information operations conducted by
8 governments and other entities [2,3]. Its intended effects include political influence,
9 increasing group polarisation, reducing trust, and generally undermining civil society
10 [4]. Effects are not limited to online processes. They regularly spill over into other
11 parts of our lives. Experimental work has shown that exposure to disinformation can
12 lead to attitude change [5] and there are many real-world examples of behaviours
13 that have been directly attributed to disinformation, such people as attacking
14 telecommunications masts in response to fake stories about ‘5G causing
15 coronavirus’ [6,7]. Social media disinformation is very widely used as a tool of
16 influence: computational propaganda has been described as a pervasive and
17 ubiquitous part of modern everyday life [8].

18 How does social media disinformation spread?

19 Once disinformation has initially been seeded online by its creators, one of the
20 ways in which it spreads is through the actions of individual social media users.
21 Ordinary people may propagate the material to their own social networks through
22 deliberate sharing – a core function of platforms such as Facebook and Twitter.
23 Other interactions with it, such as ‘liking’, also trigger the algorithms of social media
24 platforms to display it to other users. This is a phenomenon known as ‘organic reach’

1 [9]. It can lead to false information spreading exponentially. As an example, analysis
2 of the activity of the Russian 'Internet Research Agency' (IRA) disinformation group
3 in the USA between 2015 and 2017 concluded that over 30 million users shared and
4 otherwise interacted with the IRA's Facebook and Instagram posts, propagating
5 them to their families and friends [4]. There is evidence that false material is spread
6 widely and rapidly through social media due to such human behaviour [10].

7 **Why do people spread social media disinformation?**

8 When individuals share or interact with disinformation they see online, they have
9 essentially been persuaded to do so by its originators. Influential models of social
10 information processing suggest there are different routes to persuasion [e.g. 11].
11 Under some circumstances, we may carefully consider the information available. At
12 other times, we make rapid decisions based on heuristics and peripheral cues. When
13 sharing information on social media occurs, it is likely to be spontaneous and rapid,
14 rather than being a considered action that people spend time deliberating over. For
15 example, there are indications of people using the interaction features of Facebook
16 in a relatively unthinking and automatic manner [12]. In such situations, a peripheral
17 route to persuasion is likely to be important [13]. Individuals' choices to share, like and
18 so on will thus be guided primarily by heuristics or contextual cues [14].

19 Three potentially important heuristics in this context are *consistency*, *consensus*
20 and *authority* [15]. These are not the only heuristics that might possibly influence
21 whether we share false material. However, in each case there is suggestive
22 empirical evidence, and apparent real-world attempts to leverage these phenomena,
23 that make them worth considering.

24 **Consistency**

1 Consistency is the extent to which sharing would be consistent with past
2 behaviours or beliefs of the individual. For example, in the USA people with a history
3 of voting Republican might be more likely to endorse and disseminate right-wing
4 messaging [16]. There is a large body of work based on the idea that people prefer
5 to behave in ways consistent with their attitudes [17]. Research has indicated that
6 social media users consider headlines consistent with their pre-existing beliefs as
7 more credible, even when explicitly flagged as being false [18]. In the context of
8 disinformation, this could make it desirable to target audiences sympathetic to the
9 message content.

10 **Consensus**

11 Consensus is the extent to which people think their behaviour would be
12 consistent with that of most other people. In the current context, it is possible that
13 seeing a message has already been shared widely might make people more likely to
14 forward it on themselves. In marketing, this influence tactic is known as 'social proof'
15 [19]. It is widely used in online commerce in attempts to persuade consumers to
16 purchase goods or services (e.g. by displaying reviews or sales rankings). The
17 feedback mechanisms of social networks can be manipulated to create an illusion of
18 such social support, and this tactic seems to have been used in the aftermath of
19 terror attacks in the UK [20].

20 Bot networks are used to spread low-credibility information on Twitter through
21 automated means. Bots have been shown to be involved in the rapid spread of
22 information, tweeting and retweeting messages many times [21]. Among humans
23 who see the messages, the high retweet counts achieved through the bot networks
24 might be interpreted as indicating that many other people agree with them. There is
25 evidence which suggests that "each amount of sharing activity by likely bots tends to

1 trigger a disproportionate amount of human engagement" [21, p.4]. Such bot activity
2 could be an attempt to exploit the consensus effect.

3 It is relatively easy to manipulate the degree of consensus or social proof
4 associated with an online post. Work by the NATO Strategic Communications Centre
5 of Excellence [22] indicated that it was very easy to purchase high levels of false
6 engagement for social media posts (e.g. sharing of posts by networks of fake
7 accounts) and that there was a significant black market for social media
8 manipulation. Thus, if boosting consensus effectively influences organic reach, then
9 it could be a useful tool for both those seeding disinformation and those seeking to
10 spread counter-messages.

11 **Authority**

12 Authority is the extent to which the communication appears to come from a
13 credible, trustworthy source [23]. Research participants have been found to report a
14 greater likelihood of propagating a social media message if it came from a
15 trustworthy source [24]. There is evidence of real-world attempts to exploit this effect.
16 In 2018, Twitter identified fraudulent accounts that simulated those of US local
17 newspapers [25], which may be trusted more than national media [26]. These may
18 have been sleeper accounts established specifically for the purpose of building trust
19 prior to later active use.

20 **Factors influencing the spread of disinformation**

21 While there are likely to be a number of other variables that also influence the
22 spread of disinformation, there are grounds for believing that consistency, consensus
23 and authority may be important. Constructing or targeting disinformation messages
24 in such a way as to maximise these three characteristics may be a way to increase
25 their organic reach. There is real-world evidence of activity consistent with attempts

1 to exploit them. If these effects do exist, they could also be exploited by initiatives to
2 counter disinformation.

3 **Who spreads social media disinformation?**

4 Not all individuals who encounter untrue material online spread it further. In fact,
5 the great majority do not. Research linking behavioural and survey data [16] found
6 that less than 10% of participants shared articles from 'fake news' domains during
7 the 2016 US presidential election campaign (though of course when extrapolated to
8 the huge user base of social network platforms like Facebook, this is still a very large
9 number of people).

10 The fact that only a minority of people actually propagate disinformation makes it
11 important to consider what sets them apart from people who don't spread untrue
12 material further. This will help to inform interventions aimed at countering
13 disinformation. For example, those most likely to be misled by disinformation, or to
14 spread it further, could be targeted with counter-messaging. It is known that the
15 originators of disinformation have already targeted specific demographic groups, in
16 the same way as political campaigns micro-target messaging at those audience
17 segments deemed most likely to be persuadable [27]. For example, it is believed that
18 the 'Internet Research Agency' sought to segment Facebook and Instagram users
19 based on race, ethnicity and identity by targeting their messaging to people recorded
20 by the platforms as having certain interests for marketing purposes [4]. They
21 targeted specific communications tailored to those segments (e.g. trying to
22 undermine African Americans' faith in political processes and suppress their voting in
23 the US presidential election).

24 **Digital media literacy**

1 Research has found that older adults, especially those aged over 65, were by far
2 the most likely to spread material originally published by ‘fake news’ domains [16]. A
3 key hypothesis advanced to explain this is that older adults have lower levels of
4 digital media literacy, and are thus less likely to be able to distinguish between true
5 and false information online. While definitions may vary, digital media literacy can be
6 thought of as including “... the ability to interact with textual, sound, image, video and
7 social medias ... finding, manipulating and using such information” [28, p. 11] and
8 being a “multidimensional concept that comprised technical, cognitive, motoric,
9 sociological, and emotional aspects” [29, p.834]. Digital media literacy is widely
10 regarded as an important variable mediating the spread and impact of disinformation
11 [e.g. 1]. It is argued that many people lack the sophistication to detect a message as
12 being untruthful, particularly when it appears to come from an authoritative or trusted
13 source. Furthermore, people higher in digital media literacy may be more likely to
14 engage in elaborated, rather than heuristic-driven, processing (cf. work on phishing
15 susceptibility [30]), and thus be less susceptible to biases such as consistency,
16 consensus and authority.

17 Educating people in digital media literacy is the foundation of many anti-
18 disinformation initiatives. Examples include the ‘News Hero’ Facebook game
19 developed by the NATO Strategic Communications Centre of Excellence
20 (<https://www.stratcomcoe.org/news-hero>), government initiatives in Croatia and
21 France [8] or the work of numerous fact-checking organisations. The effectiveness of
22 such initiatives relies on two assumptions being met. The first is that lower digital
23 media literacy really does reduce our capacity to identify disinformation. There is
24 currently limited empirical evidence on this point, complicated by the fact that
25 definitions of ‘digital literacy’ are varied and contested, and there are currently no

1 widely accepted measurement tools [28]. The second is that the people sharing
2 disinformation are doing so unwittingly, having been tricked into spreading it.
3 However, it is possible that at least some people *know* the material is untrue, and
4 they spread it anyway. Survey research [31] has found that believing a story was
5 false was not necessarily a barrier to sharing it. People may act like this because
6 they are sympathetic to a story's intentions or message, or they are explicitly
7 signalling their social identity or allegiance to some political group or movement. If
8 people are deliberately forwarding information that they know is untrue, then raising
9 their digital media literacy would be ineffective as a stratagem to counter
10 disinformation. This makes it important to simultaneously consider users' beliefs
11 about the veracity of disinformation stories, to inform the design of countermeasures.

12 **Personality**

13 It is also known that personality influences how people use social media [e.g.
14 32]. This makes it possible that personality variables will also influence interactions
15 with disinformation. Indeed, previous research [24] found that people low on
16 Agreeableness reported themselves as more likely to propagate a message. This is
17 an important possibility to consider, because it raises the prospect that individuals
18 could be targeted on the basis of their personality traits with either disinformation or
19 counter-messaging. In a social media context, personality-based targeting of
20 communications is feasible because personality characteristics can be detected from
21 individuals' social media footprints [33, 34]. Large scale field experiments have
22 shown that personality-targeted advertising on social media can influence user
23 behaviour [35].

24 The question of which personality traits might be important is an open one. In the
25 current study, personality was approached on an exploratory basis, with no specific

1 hypotheses about effects or their directions. This is because there are a number of
2 different and potentially rival effects that might operate. For example, higher levels of
3 Conscientiousness may be associated with a greater likelihood of posting political
4 material in social media [36] leading to a higher level of political disinformation being
5 shared. However, people higher in Conscientiousness are likely to be more cautious
6 [37] and pay more attention to details [38]. They might therefore also be more likely
7 to check the veracity of the material they share, leading to a lower level of political
8 disinformation being shared.

9 **Research aims and hypotheses**

10 The overall aim of this project was to establish whether contextual factors in the
11 presentation of disinformation, or characteristics of the people seeing it, make it
12 more likely that they extend its reach. The methodology adopted was scenario-
13 based, with individuals being asked to rate their likelihood of sharing exemplar
14 disinformation messages. A series of four studies was conducted, all using the same
15 methodology. Multiple studies were used to establish whether the same effects were
16 found across different social media platforms (Facebook in Study 1, Twitter in Study
17 2, Instagram in Study 3) and countries (Facebook with a UK sample in Study 1,
18 Facebook with a US sample in Study 4). Data were also collected on whether
19 participants had shared disinformation in the past. A number of distinct hypotheses
20 were advanced:

21 H1: Individuals will report themselves as more likely to propagate messages
22 from more authoritative compared to less authoritative sources.

23 H2: Individuals will report themselves as more likely to propagate messages
24 showing a higher degree of consensus compared to those showing a lower degree
25 of consensus.

1 H3: Individuals will report themselves as more likely to propagate messages
2 consistent with their pre-existing beliefs compared to inconsistent messages.

3 H4: Individuals lower in digital literacy will report a higher likelihood of sharing
4 false messages than individuals higher in digital literacy.

5 Other variables were included in the analysis on an exploratory basis with no
6 specific hypotheses being advanced. In summary, this project asks why ordinary
7 social media users share political disinformation messages they see online. It tests
8 whether specific characteristics of messages or their recipients influence the
9 likelihood of disinformation being further shared online. Understanding any such
10 mechanisms will both increase our understanding of the phenomenon and inform the
11 design of interventions seeking to reduce its impact.

12 **Study 1**

13 Study 1 tested hypotheses 1-4 with a UK sample, using stimuli relevant to the
14 UK. The study was completed online. Participants were members of research panels
15 sourced through the research company Qualtrics.

16 **Method**

17 Participants were asked to rate their likelihood of sharing three simulated
18 Facebook posts. The study used an experimental design, manipulating levels of
19 authoritativeness and consensus apparent in the stimuli. All manipulations were
20 between, not within, participants. Consistency with pre-existing beliefs was not
21 manipulated. Instead, the political orientation of the stimuli was held constant, and
22 participants' scores on conservative political orientation were used as an index of
23 consistency between messages and participant beliefs. The effects of these
24 variables on self-rated likelihood of sharing the stimuli, along with those of a number

1 of other predictors, were assessed using multiple regression. The primary goal of the
2 analysis was to identify variables that statistically significantly explained variance in
3 the likelihood of sharing disinformation. The planned analysis was followed by
4 supplementary and exploratory analyses. All analyses were conducted using SPSS
5 v.25 for Mac. For all studies reported in this paper, ethical approval came from both
6 the University of Westminster Research Ethics Committee (ETH1819-1420) and the
7 Lancaster University Security Research Ethics Committee (BUCHANAN 2019 07
8 23). Consent was obtained, via an electronic form, from anonymous participants.

9 **Materials**

10 A short questionnaire was used to capture demographic information (gender;
11 country of residence; education; age; occupational status; political orientation
12 expressed as right, left or centre; frequency of Facebook use). Individual differences
13 in personality, political orientation, and digital / new media literacy were measured
14 using established validated questionnaires. Ecologically valid stimuli were used, with
15 their presentation being modified across conditions to vary authoritativeness and
16 consensus markers.

17 Personality was measured using a 41-item Five-Factor personality questionnaire
18 [38] derived from the International Personality Item Pool [37]. The measure provides
19 indices of Extraversion, Neuroticism, Openness to Experience, Agreeableness and
20 Conscientiousness that correlate well with the domains of Costa and McCrae's [39]
21 Five Factor Model.

22 Conservatism was measured using the 12-item Social and Economic
23 Conservatism Scale (SECS) [40], which is designed to measure political orientation
24 along a left-right; liberal-conservative continuum. It was developed and validated
25 using a US sample. In pilot work for the current study, mean scores for individuals

1 who reported voting for the Labour and Conservative parties in the 2017 UK general
2 election were found to differ in the expected manner ($t_{(28)}=-2.277$, $p=.031$, $d=0.834$).
3 This provides evidence of its appropriateness for use in UK samples. While the
4 measure provides indices of different aspects of conservatism, it also provides an
5 overall conservatism score which was used in this study.

6 Digital media literacy was measured using the 35-item New Media Literacy Scale
7 (NMLS) [29]. This is a theory-based self-report measure of competences in using,
8 critically interrogating, and creating digital media technologies and messaging. In
9 pilot work with a UK sample, it was found to distinguish between individuals high or
10 low in social media (Twitter) use, providing evidence of validity ($t_{(194)}=-3.847$, $p<.001$,
11 $d=.55$). While the measure provides indices of different aspects of new media
12 literacy, it also provides an overall score which was used in this study.

13 Participants were asked to rate their likelihood of sharing three genuine
14 examples of 'fake news' that had been previously published online. An overall score
15 for their likelihood of sharing the stimuli was obtained by summing the three ratings,
16 creating a combined score. This was done, and a set of three stimuli was used, to
17 reduce the likelihood that any effects found were peculiar to a specific story. The
18 stimuli were sourced from the website Infowars.com (which in some cases had
19 republished them from other sources). Infowars.com has been described [41] as a
20 high-exposure site strongly associated with the distribution of 'fake news'. Rather
21 than full articles, excerpts (screenshots) were used that had the size and general
22 appearance of what respondents might expect to see on social media sites. The
23 excerpts were edited to remove any indicators of the source, metrics such as the
24 numbers of shares, date, and author. All had a right-wing orientation (so that
25 participant conservatism could be used as a proxy for consistency between the

1 messages and existing beliefs). This was established in pilot work rating their
2 political orientation and likelihood of being shared. The three stories were among
3 seven rated by a UK sample ($N=30$) on an 11-point scale asking “To what extent do
4 you think this post was designed to appeal to people with right wing (politically
5 conservative) views?” anchored at “Very left wing oriented” and “Very right wing
6 oriented”. All seven were rated statistically significantly above the politically-neutral
7 midpoint of the scale. Of the three stimuli selected for use in this study, a one-
8 sample t -test showed that the least right-wing was statistically significantly higher
9 than the midpoint, ($t_{(39)}=4.385$, $p<.001$, $d=0.70$).

10 One of the stimuli was a picture of masked and hooded men titled “Censored
11 video: watch Muslims attack men, women & children in England”. One was a picture
12 of many people walking down a road, titled “Revealed: UN plan to flood America with
13 600 million migrants”, with accompanying text describing a plan to “flood America
14 and Europe with hundreds of millions of migrants to maintain population levels”. The
15 third was a picture of the Swedish flag titled “‘Child refugee’ with flagship Samsung
16 phone and gold watch complains about Swedish benefits rules”, allegedly describing
17 a 19 year-old refugee’s complaints.

18 The authoritativeness manipulation was achieved by pairing the stimuli with
19 sources regarded as relatively high or low in authoritativeness. The source was
20 shown above the stimulus being rated, in the same way as the avatar and username
21 of someone who had posted a message would be on Facebook. The lower
22 authoritativeness group were slight variants on real usernames that had previously
23 retweeted either stories from Infowars.com or another story known to be untrue. The
24 original avatars were used. The exemplars used in this study were named ‘Tigre’
25 (with an avatar of an indistinct picture of a female face), ‘jelly beans’ (a picture of

1 some jelly beans) and 'ChuckE' (an indistinct picture of a male face). The higher
2 authoritativeness group comprised actual fake accounts set up by the Internet
3 Research Agency (IRA) group to resemble local news sources, selected from a list of
4 suspended IRA accounts released by Twitter. The exemplars used in this study were
5 'Los Angeles Daily', 'Chicago Daily News' and 'El Paso Top News'. Pilot work was
6 conducted with a sample of UK participants ($N=30$) who each rated a selection of 9
7 usernames, including these 6, for the extent to which each was "likely to be an
8 authoritative source - that is, likely to be a credible and reliable source of
9 information". A within-subjects t -test indicated that mean authoritativeness ratings for
10 the 'higher' group were statistically significantly higher than the 'lower' group ($t_{(29)}=-$
11 11.181, $p<.001$, $d_z=2.04$).

12 The consensus manipulation was achieved by pairing the stimuli with indicators
13 of the number of shares and likes the story had. The indicators were shown below
14 the stimulus being rated, in the same way as they normally would be on Facebook.
15 In the low consensus conditions, low numbers of likes (1, 3, 2) and shares (2, 0, 2)
16 were displayed. In the high consensus conditions, higher (but not unrealistic)
17 numbers of likes (104K, 110K, 63K) and shares (65K, 78K, 95K) were displayed.
18 The information was presented using the same graphical indicators as would be the
19 case on Facebook, accompanied by the (inactive) icons for interacting with the post,
20 in order to maximise ecological validity.

21 **Procedure**

22 The study was conducted completely online, using materials hosted on the
23 Qualtrics research platform. Participants initially saw an information page about the
24 study, and on indicating their consent proceeded to the demographic items. They
25 then completed the personality, conservatism and new media literacy scales. Each

1 of these was presented on a separate page, except the NMLS which was split
2 across three pages.

3 Participants were then asked to rate the three disinformation items. Participants
4 were randomized to different combinations of source and story within their assigned
5 condition. For example, Participant A might have seen Story 1 attributed to Source 1,
6 Story 2 attributed to Source 2, and Story 3 attributed to Source 3; while Participant B
7 saw Story 1 attributed to Source 2, Story 2 attributed to Source 1, and Story 3
8 attributed to Source 3. Each participant saw the same three stories paired with one
9 combination of authoritativeness and consensus. There were 24 distinct sets of
10 stimuli.

11 Each participant saw an introductory paragraph stating “A friend of yours
12 recently shared this on Facebook, commenting that they thought it was important
13 and asking all their friends to share it:”. Below this was the combination of source,
14 story, and consensus indicators, presented together in the same way as a genuine
15 Facebook post would be. They then rated the likelihood of them sharing the post to
16 their own public timeline, on an 11-point scale anchored at ‘Very Unlikely’ and ‘Very
17 Likely’. This was repeated for the second and third stimuli, each on a separate page.
18 Having rated each one, participants were then shown all three stimuli again, this time
19 on the same page. They were asked to rate each one for “how likely do you think it is
20 that the message is accurate and truthful” and “how likely do you think it is that you
21 have seen it before today”, on 5-point scales anchored at ‘Not at all likely’ and ‘Very
22 likely’.

23 After rating the stimuli, participants were asked two further questions: “Have you
24 ever shared a political news story online that you later found out was made up?”, and
25 “And have you ever shared a political news story online that you thought AT THE

1 TIME was made up?”, with ‘yes’ or ‘no’ response options. This question format
2 directly replicated that used in Pew Research Centre surveys dealing with
3 disinformation [e.g. 31].

4 Finally, participants were given the opportunity once again to give or withdraw
5 their consent for participation. They then proceeded to a debriefing page. It was only
6 at the debriefing stage that they were told the stories they had seen were untrue: no
7 information about whether the stimuli were true or false had been presented prior to
8 that point.

9 **Data screening and processing**

10 Prior to delivery of the sample, Qualtrics performed a series of quality checks
11 and ‘data scrubbing’ procedures to remove and replace participants with response
12 patterns suggesting inauthentic or inattentive responding. These included speeding
13 checks and examination of response patterns. On delivery of the initial sample
14 ($N=688$) further screening procedures were performed. Sixteen respondents were
15 identified who had responded with the same scores to substantive sections of the
16 questionnaire (‘straightlining’). These were removed, leaving $N=672$. These checks
17 and exclusions were carried out prior to any data analysis. Where participants had
18 missing data on any variables, they were omitted only from analyses including those
19 variables. Thus, N s vary slightly throughout the analyses.

20 **Participants**

21 The target sample size was planned to exceed $N=614$, which would give 95%
22 power to detect $R^2 = .04$ (a benchmark for the minimum effect size likely to have real-
23 world importance in social science research [42]), in the planned multiple regression
24 analysis with 11 predictors. Qualtrics was contracted to provide a sample of
25 Facebook users that was broadly representative of the UK 2011 census population

1 in terms of gender; the split between those who had post-secondary-school
 2 education and those who had not; and age profile (18+). Quotas were used to
 3 assemble a sample comprising approximately one third each self-describing as left-
 4 wing, centre and right-wing in their political orientation. Participant demographics are
 5 shown in Table 1, column 1.

6 **Table 1. Demographic Data, Studies 1-4.**

	Study 1 (UK, N=672)	Study 2 (UK, N=674)	Study 3 (UK, N=650)	Study 4 (US, N=638)
Gender				
Male	317	312	225	283
Female	353	359	422	354
Other	1 (0.1%)	2 (0.3%)	3 (0.5%)	1 (0.2%)
Prefer not to say	1 (0.1%)	1 (0.1%)	0 (0%)	0 (0%)
Age group				
18-29	138	184	291	143
30-39	127	173	212	120
40-49	129	176	81	106
50-59	130	99	49 (7.5%)	96
60-69	96	37 (5.5%)	15 (2.3%)	160
70+	49 (7.3%)	5 (0.7%)	1 (0.2%)	12 (1.9%)
Unanswered	3 (0.4%)	0 (0%)	1 (0.2%)	1 (0.2%)
Facebook use				
Every few weeks	21 (3.1%)	-	-	21 (3.3%)
A few times a week	77	-	-	52 (8.2%)
About once a day	124	-	-	128
Several times a day	450	-	-	437
Twitter use				
Every few weeks		53 (7.9%)	-	-
A few times a week		165	-	-
About once a day		151	-	-
Several times a day		305	-	-
Instagram use				
Every few weeks	-	-	22 (3.4%)	-
A few times a week	-	-	84	-
About once a day	-	-	145	-
Several times a day	-	-	399	-

Political ideology

Left	228	247	228	199
Centre	227	237	255	229
Right	216	190	167	207
Unanswered	1 (0.2%)	0 (0%)	0 (0%)	3 (0.5%)

Highest education

Less than high school	11 (1.6%)	1 (0.1%)	2 (0.3%)	12 (1.9%)
High school / secondary	394	113	119	228
Some college or University	75	158	167	144
College or University	142	282	262	179
Master's degree	38 (5.7%)	93	82	45 (7.1%)
Doctoral degree	6 (0.9%)	16 (2.3%)	8 (1.2%)	11 (1.7%)
Professional degree (JD,	5 (0.7%)	11 (1.6%)	10 (1.5%)	19 (3.0%)
Unanswered	1 (0.1%)	0 (0%)	0 (0%)	0 (0%)

Occupation

Employed for Wages	368	430	379	312
Self-employed	39 (5.8%)	77	64 (9.8%)	55 (8.6%)
Unemployed but looking	33 (4.9%)	28 (4.2%)	27 (4.2%)	36 (5.6%)
Home-maker	70	44 (6.5%)	39 (6.0%)	45 (7.1%)
Student	26 (3.9%)	60 (8.9%)	113	23 (3.6%)
Retired	96	23 (3.4%)	10 (1.5%)	121
Unable to work for health	39 (5.8%)	12 (1.8%)	18 (2.8%)	46 (7.2%)
Unanswered	1 (0.1%)	0 (0%)	0 (0%)	0 (0%)

1

2

3 Results

4 Descriptive statistics for participant characteristics (personality, conservatism, new
5 media literacy and age) and their reactions to the stimuli (likelihood of sharing, belief
6 the stories were likely to be true, and rating of likelihood that they had seen them
7 before) are summarised in Table 2. All scales had acceptable reliability. The main
8 dependent variable, likelihood of sharing, had a very skewed distribution with a
9 strong floor effect: 39.4% of the participants indicated they were 'very unlikely' to
10 share any of the three stories they saw. This is consistent with findings on real-world
11 sharing that indicate only a small proportion of social media users will actually share

1 disinformation [e.g. 16], though it gives a dependent variable with less than ideal
2 distributional properties.

3 **Table 2. Descriptive Statistics for Participant Characteristics and Response**

4 **Variables, Study 1.**

	<i>N</i>	<i>M</i>	<i>SD</i>	α	Range		Skew	Kurtosis
					Potential	Actual		
Openness	668	22.55	5.22	0.70	7-35	8-35	-0.20	-0.16
Conscientiousness	667	36.03	6.10	0.79	10-50	12-50	-0.14	-0.04
Extraversion	669	26.56	6.33	0.81	9-45	9-44	-0.19	0.10
Agreeableness	669	26.85	4.49	0.78	7-35	13-35	-0.52	-0.16
Neuroticism	669	21.97	7.02	0.87	8-40	8-40	0.14	-0.39
Conservatism	670	651.00	146.26	0.72	0-1200	160-1090	-0.43	0.12
New Media Literacy	652	124.76	20.18	0.95	35-175	55-172	-0.26	0.01
Likelihood of sharing	672	8.92	7.49		3-33	3-33	1.32	0.97
Belief stories true	672	6.82	3.18		3-15	3-15	0.60	-0.40
Likelihood seen before	672	6.10	3.34		3-15	3-15	0.83	-0.40
Age	669	45.32	15.91			18-86	0.13	-0.94

5

6

7 To simultaneously test hypotheses 1-4 a multiple regression analysis was
8 carried out. This evaluated the extent to which digital media literacy (NMLS),
9 authority of the message source, consensus, belief in veracity of the messages,
10 consistency with participant beliefs (operationalised as the total SECS conservatism
11 scale score), age and personality (Extraversion, Conscientiousness, Agreeableness,
12 Openness to Experience and Neuroticism), predicted self-rated likelihood of sharing

1 the posts. This analysis is summarised in Table 3. Checks were performed on
 2 whether the dataset met the assumptions required by the analysis (absence of
 3 collinearity, independence of residuals, heteroscedasticity and non-normal
 4 distribution of residuals). Despite the skewed distribution of the dependent variable,
 5 no significant issues were detected.

6 **Table 3. Predictors of Rated Likelihood of Sharing Stimuli, Study 1.**

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Constant	1.596	3.363		0.475	.635
New Media Literacy	0.026	0.013	0.069	1.908	.057
Conservatism	0.005	0.002	0.090	2.505	.013
Openness	-0.094	0.053	-0.066	-1.786	.075
Neuroticism	-0.037	0.042	-0.035	-0.885	.377
Extraversion	0.076	0.044	0.064	1.744	.082
Conscientiousness	-0.064	0.047	-0.051	-1.362	.174
Agreeableness	-0.183	0.062	-0.108	-2.947	.003
Authoritativeness	0.091	0.478	0.006	0.189	.850
Consensus	0.606	0.480	0.040	1.263	.207
Belief stories true	1.292	0.080	0.543	16.139	.000
Age	0.002	0.017	0.003	0.090	.928
<i>R</i> ²					.377
Adjusted <i>R</i> ²					.366
<i>F</i>					34.353

7 Authoritativeness and Consensus conditions coded as dummy variables;

8 higher=1, lower=0 in each case. *N*=636.

9

1 However, exploratory analyses indicated that inclusion of other variables in the
2 regression model might be warranted. It is well established that there are gender
3 differences on a number of personality variables. Furthermore, in the current sample
4 men and women differed in their level of conservatism ($M=669.10$, $SD=150.68$ and
5 $M=636.50$, $SD=138.31$ respectively; $t_{(666)}=2.914$, $p=.004$), their self-rated likelihood of
6 sharing ($M=10.41$, $SD=8.33$ and $M=7.60$, $SD=6.38$ respectively; $t_{(589.60)}=4.928$,
7 $p<.001$; adjusted df used due to heterogeneity of variance, Levene's $F=35.99$,
8 $p<.001$), and their belief that the stories were true ($M=7.16$, $SD=3.22$ and $M=6.52$,
9 $SD=3.12$ respectively; $t_{(668)}=2.574$, $p=.010$). Education level was found to correlate
10 positively with NMLS scores ($r=.210$, $N=651$, $p<.001$). Level of Facebook use
11 correlated significantly with age ($r=-.126$, $N=669$, $p=.001$), education ($r=.082$, $N=671$,
12 $p=.034$), NMLS ($r=.170$, $N=652$, $p<.001$), with likelihood of sharing ($r=.079$, $N=672$,
13 $p=.040$), and with likelihood of having seen the stimuli before ($r=.107$, $N=672$,
14 $p=.006$). Self-reported belief that respondents had seen the stories before also
15 correlated significantly with likelihood of sharing ($r=.420$, $N=672$, $p<.001$), and a
16 number of other predictor variables.

17 Accordingly, a further regression analysis was performed, including these
18 additional predictors (gender, education, level of Facebook use, belief they had seen
19 the stories before). Given inclusion of gender as a predictor variable, the two
20 respondents who did not report their gender as either male or female were excluded
21 from further analysis. The analysis, summarised in Table 4, indicated that the model
22 explained 43% of the variance in self-reported likelihood of sharing the three
23 disinformation items. Neither the authoritativeness of the story source, nor
24 consensus information associated with the stories, was a significant predictor.

1 Consistency of the items with participant attitudes (conservatism) was important,
 2 with a positive and statistically significant relationship between conservatism and
 3 likelihood of sharing. The only personality variable predicting sharing was
 4 Agreeableness, with less agreeable people giving higher ratings of likelihood of
 5 sharing. In terms of demographic characteristics, gender and education were
 6 statistically significant predictors, with men and less-educated people reporting a
 7 higher likelihood of sharing. Finally, people reported a greater likelihood of sharing
 8 the items if they believed they were likely to be true, and if they thought they had
 9 seen them before.

10 **Table 4. Predictors of Rated Likelihood of Sharing Stimuli (Extended Predictor**
 11 **Set), Study 1.**

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Constant	0.395	3.473		0.114	.909
New Media Literacy	0.012	.013	.031	0.861	.390
Conservatism	0.005	.002	.097	2.828	.005
Openness	-0.059	.052	-.041	-1.136	.257
Neuroticism	-0.007	.041	-.007	-0.180	.857
Extraversion	0.075	.042	.064	1.796	.073
Conscientiousness	-0.017	.045	-.014	-0.376	.707
Agreeableness	-0.123	.061	-.072	-2.023	.044
Authoritativeness	0.319	.463	.021	0.690	.490
Consensus	0.563	.462	.037	1.219	.223
Belief stories true	1.059	.084	.444	12.573	.000
Age	-0.014	.017	-.029	-0.778	.437
Gender (M=1, F=2)	-1.602	.508	-.106	-3.153	.002

Education	-0.709	.224			-.103	-3.160	.002
Facebook use	0.369	.294			.041	1.257	.209
Likelihood seen before	0.479	.079			.211	6.053	.000
R^2					.431		
Adjusted R^2					.418		
F					31.258		.000

1 Authoritativeness and Consensus conditions coded as dummy variables;
2 higher=1, lower=0 in each case. $N=634$.

3

4 Participants had also been asked about their historical sharing of untrue political
5 stories, both unknowing and deliberate. 102 out of 672 participants (15.2%) indicated
6 that they had ever 'shared a political news story online that they later found out was
7 made up', while 64 out of 672 indicated they had shared one that they 'thought AT
8 THE TIME was made up' (9.5%). Predictors of whether or not people had shared
9 untrue material under both sets of circumstances were examined using logistic
10 regressions, with the same sets of participant-level predictors.

11 Having unknowingly shared untrue material (Table 5) was significantly predicted
12 by lower Conscientiousness, lower Agreeableness, and lower age. Having shared
13 material known to be untrue at the time (Table 6) was significantly predicted by lower
14 Agreeableness and lower age.

15 **Table 5. Binary Logistic Regression: Predictors of Whether Participants had**
16 **Previously Shared Political Stories They Subsequently Discovered Were False,**
17 **Study 1.**

	B	$S.E.$	$Wald$	df	p	$Exp(B)$	95% C.I. for $EXP(B)$
New Media Literacy	0.004	0.007	0.415	1	.519	1.004	[0.991, 1.018]

Conservatism	0.001	0.001	0.339	1	.560	1.001	[0.999, 1.002]
Openness	-0.040	0.027	2.157	1	.142	0.961	[0.912, 1.013]
Neuroticism	-0.007	0.023	0.088	1	.767	0.993	[0.950, 1.039]
Extraversion	0.036	0.023	2.336	1	.126	1.036	[0.990, 1.085]
Conscientiousness	-0.053	0.024	4.971	1	.026	0.949	[0.906, 0.994]
Agreeableness	-0.085	0.030	8.26	1	.004	0.918	[0.866, 0.973]
Age	-0.024	0.009	7.416	1	.006	0.976	[0.960, 0.993]
Gender (M=1, F=2)	-0.463	0.258	3.213	1	.073	0.629	[0.379, 1.044]
Education	0.082	0.108	0.582	1	.446	1.086	[0.879, 1.341]
Facebook use	0.336	0.176	3.644	1	.056	1.400	[0.991, 1.978]
Constant	1.463	1.736	0.71	1	.399	4.320	

1 *N*=634.

2

3 **Table 6. Binary Logistic Regression: Predictors of Whether Participants had**

4 **Previously Shared Political Stories They Knew at the Time Were False, Study**

5 **1.**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp(B)</i>	<i>95% C.I. for EXP(B)</i>
New Media Literacy	0.003	0.008	0.138	1	.711	1.003	[0.987, 1.019]
Conservatism	0.000	0.001	0.001	1	.981	1.000	[0.998, 1.002]
Openness	0.013	0.037	0.117	1	.732	1.013	[0.942, 1.088]
Neuroticism	-0.001	0.030	0.002	1	.965	0.999	[0.942, 1.059]
Extraversion	0.054	0.032	2.854	1	.091	1.056	[0.991, 1.124]
Conscientiousness	-0.043	0.031	1.983	1	.159	0.958	[0.902, 1.017]
Agreeableness	-0.173	0.039	20.169	1	.000	0.841	[0.780, 0.907]
Age	-0.042	0.012	12.827	1	.000	0.959	[0.938, 0.981]

Gender (M=1, F=2)	-0.602	0.325	3.420	1	.064	0.548	[0.290, 1.037]
Education	0.123	0.128	0.918	1	.338	1.131	[0.879, 1.455]
Facebook use	0.181	0.214	0.715	1	.398	1.198	[0.788, 1.822]
Constant	2.920	2.179	1.796	1	.180	18.548	

1 *N*=634.

2

3 **Discussion**

4 The main analysis in this study (Table 4) provided limited support for the
5 hypotheses. Contrary to hypotheses 1, 2, and 4, neither consensus markers,
6 authoritativeness of source, nor new media literacy were associated with self-rated
7 likelihood of sharing the disinformation stories. However, in line with hypothesis 3,
8 higher levels of conservatism were associated with higher likelihood of sharing
9 disinformation. This finding supports the proposition that we are more likely to share
10 things that are consistent with our pre-existing beliefs, as all the stimuli were right-
11 wing in orientation. An alternative explanation might be that more conservative
12 people are simply more likely to share disinformation. However, as well as lacking a
13 solid rationale, this explanation is not supported by the fact that conservatism did not
14 seem to be associated with self-reported historical sharing (Tables 5 and 6).

15 The strongest predictors of likelihood of sharing were belief that the stories were
16 true, and likelihood of having seen them before. Belief in the truth of the stories
17 provides further evidence for the role of consistency (hypothesis 3), in that we are
18 more likely to share things we believe are true. The association with likely previous
19 exposure to the materials is consistent with other recent research [43, 44] that found
20 that prior exposure to 'fake news' headlines led to higher belief in their accuracy and
21 reduced belief that it would be unethical to share them.

1 Of the personality variables, only Agreeableness was a significant predictor, with
2 less agreeable people rating themselves are more likely to share the stimuli. This is
3 consistent with previous findings [24] that less agreeable people reported they were
4 more likely to share a critical political message.

5 Lower education levels were associated with a higher self-reported likelihood of
6 sharing. It is possible that less educated people may be more susceptible to online
7 influence, given work finding that less educated people were more influenced by
8 micro-targeted political advertising on Facebook [45].

9 Finally, gender was found to be an important variable, with men reporting a
10 higher likelihood of sharing the disinformation messages than women. This was
11 unanticipated: while there are a number of gender-related characteristics (e.g.
12 personality traits) that were thought might be important, there were no *a priori*
13 grounds to expect that gender itself would be a predictor variable.

14 Study 1 also examined predictors of reported historical sharing of false political
15 information. Consistent with real-world data [16], and past representative surveys
16 [e.g. 31], a minority of respondents reported such past sharing. Unknowingly sharing
17 false political stories was predicted by low Conscientiousness, low Agreeableness,
18 and lower age, while knowingly sharing false material was predicted only by lower
19 Agreeableness and lower age. The effect of Agreeableness is consistent with the
20 findings from the main analysis and from [24]. The finding that Conscientiousness
21 influenced accidental, but not deliberate, sharing is consistent with the idea that less
22 conscientious people are less likely to check the details or veracity of a story before
23 sharing it. Clearly this tendency would not apply to deliberate sharing of falsehoods.
24 The age effect is harder to explain, especially given evidence [16] that older people
25 were more likely to share material from fake news sites. One possible explanation is

1 that younger people are more active on social media, so would be more likely to
2 share any kind of article. Another possibility is that they are more likely to engage in
3 sharing humorous political memes, which could often be classed as false political
4 stories.

5 **Study 2**

6 Study 2 set out to repeat Study 1, but presented the materials as if they had
7 been posted on Twitter rather than Facebook. The purpose of this was to test
8 whether the observed effects applied across different platforms. Research
9 participants have reported using 'likes' on Twitter in a more considered manner than
10 on Facebook [12], raising the possibility that heuristics might be less important for
11 this platform. The study was completed online, using paid respondents sourced from
12 the Prolific research panel (www.prolific.co).

13 **Method**

14 The methodology exactly replicated that of Study 1, except in the case of details
15 noted below. The planned analysis was revised to include the expanded set of
16 predictors eventually used in Study 1 (see Table 4).

17 **Materials**

18 Measures and materials were the same as used in Study 1. The key difference
19 from Study 1 was in the presentation of the three stimuli, which were portrayed as
20 having been posted to Twitter rather than Facebook. For the authoritativeness
21 manipulation, the screen names of the sources were accompanied by @usernames,
22 as is conventional on Twitter. For the consensus manipulation, 'retweets' were
23 displayed rather than 'shares', and the appropriate icons for Twitter were used.
24 Participants also indicated their level of Twitter, rather than Facebook, use.

1 **Procedure**

2 The procedure replicated Study 1, save that in this case the NMLS was
3 presented on a single page. Before participants saw each of the three disinformation
4 items, the introductory paragraph stated “A friend of yours recently shared this on
5 Twitter, commenting that they thought it was important and asking all their friends to
6 retweet it:”, and they were asked to indicate the likelihood of them ‘retweeting’ rather
7 than ‘sharing’ the post.

8 **Data screening and processing**

9 Data submissions were initially obtained from 709 participants. A series of
10 checks were performed to ensure data quality, resulting in a number of responses
11 being excluded. One individual declined consent. Eleven were judged to have
12 responded inauthentically, with the same responses to all items in substantive
13 sections of the questionnaire (‘straightlining’). Twenty were not active Twitter users:
14 three individuals visited Twitter ‘not at all’ and seventeen ‘less often’ than every few
15 weeks. Three participants responded unrealistically quickly, with response durations
16 shorter than four minutes (the same value used as a speeding check by Qualtrics in
17 Study 1). All of these respondents were removed, leaving $N=674$. These checks and
18 exclusions were carried out prior to any data analysis.

19 **Participants**

20 The target sample size was planned to exceed $N=614$, as in Study 1. No attempt
21 was made to recruit a demographically representative sample: instead, sampling
22 quotas were used to ensure the sample was not homogenous with respect to
23 education (pre-degree vs. undergraduate degree or above), age (under 40 vs. over
24 40) and political preference (left, centre or right wing orientation). Additionally,
25 participants had to be UK nationals resident in the UK; active Twitter users; and not

1 participants in prior studies related to this one. Each participant received a reward of
 2 £1.25. Participant demographics are shown in Table 1 (column 2). For the focal
 3 analysis in this study, the sample size conferred 94.6% power to detect $R^2 = .04$ in a
 4 multiple regression with 15 predictors (2-tailed, $\alpha = .05$).

5 Results

6 Descriptive statistics are summarised in Table 7. All scales had acceptable
 7 reliability. The main dependent variable, likelihood of sharing, again had a very
 8 skewed distribution with a strong floor effect.

9 **Table 7. Descriptive Statistics for Participant Characteristics and Response**

10 **Variables, Study 2.**

	<i>N</i>	<i>M</i>	<i>SD</i>	α	Range		Skew	Kurtosis
					Potential	Actual		
Openness	674	24.42	5.68	.76	7-35	9-35	-0.20	-0.54
Conscientiousness	674	35.85	6.42	.84	10-50	12-50	-0.35	0.06
Extraversion	674	26.24	7.05	.88	9-45	9-45	-0.01	-0.42
Agreeableness	674	27.33	3.94	.76	7-35	13-35	-0.51	0.19
Neuroticism	674	21.66	7.00	.88	8-40	8-40	0.13	-0.68
Conservatism	674	612.72	173.30	.82	0-1200	140-1150	-0.11	-0.21
New Media Literacy	674	134.55	17.05	.93	35-175	59-173	-0.14	0.09
Likelihood of retweeting	674	6.61	5.94		3-33	3-33	1.98	3.54
Belief stories true	674	5.87	2.68		3-15	3-15	0.86	0.21
Likelihood seen before	674	5.70	3.00		3-15	3-15	1.06	0.43
Age	674	38.85	12.64			18-75	0.37	-0.59

11

12

1 To simultaneously test hypotheses 1-4, a multiple regression analysis was
 2 carried out using the expanded predictor set from Study 1. Given inclusion of gender
 3 as a predictor variable, the three respondents who did not report their gender as
 4 either male or female were excluded from further analysis. The analysis,
 5 summarised in Table 8, indicated that the model explained 46% of the variance in
 6 self-reported likelihood of sharing the three disinformation items. Neither the
 7 authoritativeness of the story source, nor consensus information associated with the
 8 stories, nor new media literacy, was a significant predictor. Consistency of the items
 9 with participant attitudes (conservatism) was important, with a positive and
 10 statistically significant relationship between conservatism and likelihood of sharing.
 11 No personality variable predicted ratings of likelihood of sharing. In terms of
 12 demographic characteristics, gender and education were statistically significant
 13 predictors, with men and less-educated people reporting a higher likelihood of
 14 sharing. Finally, people reported a greater likelihood of sharing the items if they
 15 believed they were likely to be true, and if they thought they had seen them before.

16 **Table 8. Predictors of Rated Likelihood of Retweeting Stimuli (Extended**
 17 **Predictor Set), Study 2.**

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Constant	-3.027	2.717		-1.114	.266
New Media Literacy	0.005	0.011	0.013	0.421	.674
Conservatism	0.004	0.001	0.107	2.788	.005
Openness	0.021	0.037	0.020	0.581	.561
Neuroticism	0.013	0.030	0.015	0.441	.659
Extraversion	0.018	0.027	0.021	0.663	.508
Conscientiousness	0.038	0.031	0.042	1.231	.219

Agreeableness	-0.070	0.048	-0.047	-1.449	.148
Authoritativeness	-0.402	0.345	-0.034	-1.165	.244
Consensus	0.354	0.343	0.030	1.032	.302
Belief stories true	1.238	0.074	0.557	16.742	.000
Age	0.015	0.015	0.032	1.031	.303
Gender (M=1, F=2)	-1.026	0.368	-0.086	-2.789	.005
Education	-0.375	0.162	-0.069	-2.316	.021
Twitter use	0.017	0.177	0.003	0.095	.925
Likelihood seen before	0.191	0.059	0.096	3.226	.001
<i>R</i> ²				.459	
Adjusted <i>R</i> ²				.447	
<i>F</i>				37.070	.000

1 Authoritativeness and Consensus conditions coded as dummy variables;
2 higher=1, lower=0 in each case. *N*=671.

3

4 Participants had also been asked about their historical sharing of untrue political
5 stories, both unknowing and deliberate. 102 out of 674 participants (15.1%) indicated
6 that they had out ever 'shared a political news story online that they later found out
7 was made up', while 42 out of 674 indicated they had shared one that they 'thought
8 AT THE TIME was made up' (6.2%). Predictors of whether or not people had shared
9 untrue material under both sets of circumstances were examined using logistic
10 regressions, with the same sets of participant-level predictors.

11 Having unknowingly shared untrue material (Table 9) was significantly predicted
12 by higher Extraversion and higher levels of Twitter use. Having shared material

1 known to be untrue at the time (Table 10) was significantly predicted by higher
 2 Neuroticism and being male.

3 **Table 9. Binary Logistic Regression: Predictors of Whether Participants had**
 4 **Previously Shared Political Stories They Subsequently Discovered Were False,**
 5 **Study 2.**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp(B)</i>	<i>95% C.I. for EXP(B)</i>
New Media Literacy	0.002	0.007	0.055	1	.815	1.002	[0.988, 1.016]
Conservatism	0.001	0.001	3.038	1	.081	1.001	[1.000, 1.003]
Openness	0.039	0.024	2.596	1	.107	1.040	[0.992, 1.091]
Neuroticism	0.027	0.020	1.853	1	.173	1.027	[0.988, 1.067]
Extraversion	0.040	0.018	5.299	1	.021	1.041	[1.006, 1.078]
Conscientiousness	0.003	0.020	0.025	1	.874	1.003	[0.964, 1.044]
Agreeableness	0.005	0.031	0.023	1	.880	1.005	[0.946, 1.067]
Age	-0.015	0.010	2.481	1	.115	0.985	[0.966, 1.004]
Gender (M=1, F=2)	-0.228	0.240	0.908	1	.341	0.796	[0.498, 1.273]
Education	0.094	0.105	0.806	1	.369	1.099	[0.895, 1.349]
Twitter use	0.356	0.126	7.997	1	.005	1.427	[1.115, 1.826]
Constant	-6.619	1.779	13.847	1	.000	0.001	

6 *N*=671.

7

8 **Table 10. Binary Logistic Regression: Predictors of Whether Participants had**
 9 **Previously Shared Political Stories They Knew at the Time Were False, Study**
 10 **2.**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp(B)</i>	<i>95% C.I. for EXP(B)</i>
New Media Literacy	0.009	0.011	0.733	1	.392	1.009	[0.988, 1.030]

Conservatism	0.001	0.001	1.255	1	.263	1.001	[0.999, 1.004]
Openness	0.030	0.036	0.667	1	.414	1.030	[0.959, 1.106]
Neuroticism	0.065	0.029	5.018	1	.025	1.067	[1.008, 1.129]
Extraversion	0.034	0.026	1.721	1	.190	1.034	[0.984, 1.087]
Conscientiousness	0.006	0.030	0.042	1	.837	1.006	[0.949, 1.067]
Agreeableness	-0.049	0.044	1.252	1	.263	0.952	[0.874, 1.037]
Age	-0.017	0.014	1.466	1	.226	0.983	[0.955, 1.011]
Gender (M=1, F=2)	-1.093	0.384	8.093	1	.004	0.335	[0.158, 0.712]
Education	-0.205	0.165	1.553	1	.213	0.815	[0.590, 1.125]
Twitter use	0.065	0.178	0.134	1	.714	1.067	[0.753, 1.514]
Constant	-4.125	2.583	2.551	1	.110	0.016	

1 *N*=671.

2

3 **Discussion**

4 For the main analysis, Study 2 replicates a number of key findings from Study 1.

5 In particular, hypotheses 1, 2 and 4 were again unsupported by the results:

6 consensus, authoritativeness, and new media literacy were not associated with self-

7 rated likelihood of retweeting the disinformation stories. Evidence consistent with

8 hypothesis 3 was again found, with higher levels of conservatism being associated

9 with higher likelihood of retweeting. Again, the strongest predictor of likelihood of

10 sharing was belief that the stories were true, while likelihood of having seen them

11 before was again statistically significant. The only difference was in the role of

12 personality: there was no association between Agreeableness (or any other

13 personality variable) and likelihood of retweeting the material.

1 However, for self-reports of historical sharing of false political stories, the pattern
2 of results was different. None of the previous results were replicated, and new
3 predictors were observed for both un-knowing and deliberate sharing. For
4 unintentional sharing, the link with higher levels of Twitter use makes sense, as
5 higher usage confers more opportunities to accidentally share untruths. Higher
6 Extraversion has also been found to correlate with higher levels of social media use
7 [32] so the same logic may apply for that variable. For intentional sharing, the finding
8 that men were more likely to share false political information is similar to findings
9 from Study 1. The link with higher Neuroticism is less easy to explain: one possibility
10 is that more neurotic people are more likely to share falsehoods that will reduce the
11 chances of an event that they worry about (for example, spreading untruths about a
12 political candidate who one is worried about being elected).

13 Given that these questions asked about past behaviour in general, and were not
14 tied to the Twitter stimuli used in this study, it is not clear why the pattern of results
15 should have differed from those in Study 1. One possibility is that the sample
16 characteristics were different (this sample was younger, better educated, and drawn
17 from a different source). Another realistic possibility, especially given the typically low
18 effect sizes and large samples tested, is that these are simply ‘crud’ correlations [46]
19 rather than useful findings. Going forward, it is likely to be more informative to focus
20 on results that replicate across multiple studies or conceptually similar analyses.

21 **Study 3**

22 Study 3 set out to repeat Study 1, but presented the materials as if they had
23 been posted on Instagram rather than Facebook. Instagram presents an interesting
24 contrast, as the mechanisms of engagement with material are different (for example
25 there is no native sharing mechanism). Nonetheless, it has been identified as an

1 important theater for disinformation operations [47]. Study 3 therefore sought to
2 establish whether the same factors affecting sharing on Facebook also affect
3 engagement with false material on Instagram. The study was completed online,
4 using paid respondents sourced from the Prolific research panel.

5 **Method**

6 The methodology exactly replicated that of Study 1, except in the case of details
7 noted below. The planned analysis was revised to include the expanded set of
8 predictors eventually used in Study 1 (see Table 4).

9 **Materials**

10 Measures and materials were the same as used in Study 1. The only difference
11 from Study 1 was in the presentation of the three stimuli, which were portrayed as
12 having been posted to Instagram rather than Facebook. For the consensus
13 manipulation, 'likes' were used as the sole consensus indicator, and the appropriate
14 icons for Instagram were used.

15 **Procedure**

16 The procedure replicated Study 1, save that in this case the NMLS was
17 presented on a single page. Before participants saw each of the three disinformation
18 items, the introductory paragraph stated "Imagine that you saw this post on your
19 Instagram feed:" and they were asked to indicate the probability of them 'liking' the
20 post.

21 **Data screening and processing**

22 Data submissions were initially obtained from 692 participants. A series of
23 checks were performed to ensure data quality, resulting in a number of responses
24 being excluded. Four individuals declined consent. Twenty-one were judged to have

1 responded inauthentically, with the same scores to substantive sections of the
2 questionnaire ('straightlining'). Five did not indicate they were located in the UK. Ten
3 were not active Instagram users: three individuals visited Instagram 'not at all' and
4 seven 'less often' than every few weeks. Two participants responded unrealistically
5 quickly, with response durations shorter than four minutes (the same value used as a
6 speeding check by Qualtrics in Study 1). All of these respondents were removed,
7 leaving $N=650$. These checks and exclusions were carried out prior to any data
8 analysis.

9 **Participants**

10 The target sample size was planned to exceed $N=614$, as in Study 1. No attempt
11 was made to recruit a demographically representative sample: instead, sampling
12 quotas were used to ensure the sample was not homogenous with respect to
13 education (pre-degree vs. undergraduate degree or above) and political preference
14 (left, centre or right-wing orientation). Sampling was not stratified by age, given that
15 Instagram use is associated with younger ages, and the number of older Instagram
16 users in the Prolific pool was limited at the time the study was carried out.
17 Additionally, participants had to be UK nationals resident in the UK; active Instagram
18 users; and not participants in prior studies related to this one. Each participant
19 received a reward of £1.25. Participant demographics are shown in Table 1 (column
20 3). For the focal analysis in this study, the sample size conferred 93.6% power to
21 detect $R^2 = .04$ in a multiple regression with 15 predictors (2-tailed, $\alpha = .05$).

22 **Results**

23 Descriptive statistics are summarised in Table 11. All scales had acceptable
24 reliability. The main dependent variable, probability of liking, again had a very
25 skewed distribution with a strong floor effect.

1 **Table 11. Descriptive Statistics for Participant Characteristics and Response**
 2 **Variables, Study 3.**

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>α</i>	Range		Skew	Kurtosis
					Potential	Actual		
Openness	650	24.27	5.71	.75	7-35	8-35	-0.29	-0.29
Conscientiousness	650	35.44	6.71	.84	10-50	15-50	-0.22	-0.24
Extraversion	650	26.66	7.31	.89	9-45	9-45	-0.06	-0.44
Agreeableness	649	27.09	4.29	.77	7-35	9-35	-0.71	0.78
Neuroticism	650	23.23	7.29	.88	8-40	8-40	0.13	-0.70
Conservatism	650	614.34	166.46	.79	0-1200	120-1130	-0.34	0.26
New Media Literacy	650	134.66	16.85	.93	35-175	69-174	-0.34	0.50
Probability of liking	650	5.68	4.47		3-33	3-29	2.27	5.94
Belief stories true	650	6.01	2.65		3-15	3-15	0.77	0.12
Likelihood seen before	650	5.19	2.85		3-15	3-15	1.36	1.20
Age	649	32.73	11.12			18-75	0.90	0.37

3

4

5 To simultaneously test hypotheses 1-4, a multiple regression analysis was
 6 carried out using the expanded predictor set from Study 1. Given inclusion of gender
 7 as a predictor variable, the three respondents who did not report their gender as
 8 either male or female were excluded from further analysis. The analysis,
 9 summarised in Table 12, indicated that the model explained 24% of the variance in
 10 self-reported likelihood of sharing the three disinformation items. Neither the
 11 authoritativeness of the story source, consensus information associated with the
 12 stories, nor consistency of the items with participant attitudes (conservatism) was a

1 statistically significant predictor. Extraversion positively and Conscientiousness
 2 negatively predicted ratings of likelihood of sharing. In terms of demographic
 3 characteristics, men and younger participants reporting a higher likelihood of
 4 sharing. Finally, people reported a greater likelihood of sharing the items if they
 5 believed they were likely to be true, and if they thought they had seen them before.

6 **Table 12. Predictors of Rated Likelihood of Liking Stimuli (Extended Predictor**
 7 **Set), Study 3.**

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Constant	9.334	2.465		3.786	.000
New Media Literacy	-0.012	0.010	-0.046	-1.215	.225
Conservatism	0.001	0.001	0.044	1.016	.310
Openness	0.035	0.032	0.044	1.066	.287
Neuroticism	-0.050	0.028	-0.082	-1.795	.073
Extraversion	0.008	0.025	0.014	0.338	.736
Conscientiousness	-0.005	0.029	-0.008	-0.177	.860
Agreeableness	-0.093	0.041	-0.089	-2.282	.023
Authoritativeness	-0.517	0.315	-0.058	-1.641	.101
Consensus	-0.194	0.313	-0.022	-0.620	.535
Belief stories true	0.584	0.065	0.345	9.001	.000
Age	-0.047	0.016	-0.117	-3.024	.003
Gender (M=1, F=2)	-1.455	0.361	-0.155	-4.031	.000
Education	-0.081	0.151	-0.019	-0.537	.592
Instagram use	0.076	0.202	0.014	0.376	.707
Likelihood seen before	0.162	0.056	0.103	2.877	.004
<i>R</i> ²					.241

Adjusted R^2	.223	
F	13.342	.000

1 Authoritativeness and Consensus conditions coded as dummy variables;
2 higher=1, lower=0 in each case. $N=645$.

3

4 Participants had also been asked about their historical sharing of untrue political
5 stories, both unknowing and deliberate. Eighty five out of 650 (13.1%) participants
6 who answered the question indicated that they had out ever 'shared a political news
7 story online that they later found out was made up', while 50 out of 650 indicated
8 they had shared one that they 'thought AT THE TIME was made up' (7.7%).

9 Predictors of whether or not people had shared untrue material under both sets of
10 circumstances were examined using logistic regressions, with the same sets of
11 participant-level predictors.

12 Having unknowingly shared untrue material (Table 13) was significantly
13 predicted by higher Extraversion, lower Conscientiousness and male gender. Having
14 shared material known to be untrue at the time (Table 14) was significantly predicted
15 by higher New Media Literacy, higher Conservatism, and higher Neuroticism.

16 **Table 13. Binary Logistic Regression: Predictors of Whether Participants had**
17 **Previously Shared Political Stories They Subsequently Discovered Were False,**
18 **Study 3.**

	B	$S.E.$	$Wald$	df	p	$Exp(B)$	95% C.I. for $EXP(B)$
New Media Literacy	0.008	0.008	0.938	1	.333	1.008	[0.992, 1.023]
Conservatism	0.000	0.001	0.018	1	.893	1.000	[0.998, 1.002]
Openness	0.020	0.025	0.611	1	.434	1.020	[0.971, 1.071]
Neuroticism	0.036	0.022	2.722	1	.099	1.036	[0.993, 1.081]

Extraversion	0.057	0.019	9.404	1	.002	1.059	[1.021, 1.098]
Conscientiousness	-0.050	0.022	5.302	1	.021	0.951	[0.911, 0.993]
Agreeableness	0.004	0.030	0.018	1	.894	1.004	[0.947, 1.065]
Age	0.011	0.012	0.814	1	.367	1.011	[0.988, 1.034]
Gender (M=1, F=2)	-0.754	0.267	7.991	1	.005	0.471	[0.279, 0.794]
Education	-0.127	0.119	1.153	1	.283	0.880	[0.698, 1.111]
Instagram use	0.138	0.160	0.743	1	.389	1.148	[0.839, 1.570]
Constant	-3.402	1.856	3.360	1	.067	0.033	[0.992, 1.023]

1 N=645.

2

3 **Table 14. Binary Logistic Regression: Predictors of Whether Participants had**

4 **Previously Shared Political Stories They Knew at the Time Were False, Study**

5 **3.**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp(B)</i>	<i>95% C.I. for EXP(B)</i>
New Media Literacy	0.026	0.011	6.026	1	.014	1.026	[1.005, 1.048]
Conservatism	0.003	0.001	5.976	1	.015	1.003	[1.001, 1.005]
Openness	0.042	0.032	1.745	1	.187	1.043	[0.980, 1.111]
Neuroticism	0.075	0.028	7.372	1	.007	1.078	[1.021, 1.138]
Extraversion	0.042	0.023	3.344	1	.067	1.043	[0.997, 1.091]
Conscientiousness	-0.029	0.027	1.187	1	.276	0.971	[0.922, 1.023]
Agreeableness	-0.054	0.035	2.446	1	.118	0.947	[0.884, 1.014]
Age	-0.001	0.015	0.008	1	.927	0.999	[0.969, 1.029]
Gender (M=1, F=2)	-0.603	0.344	3.069	1	.080	0.547	[0.279, 1.074]
Education	-0.163	0.156	1.100	1	.294	0.849	[0.626, 1.152]
Instagram use	-0.191	0.191	0.997	1	.318	0.826	[0.568, 1.201]

Constant	-6.784	2.381	8.119	1	.004	0.001
----------	--------	-------	-------	---	------	-------

1 *N*=645.

2

3 **Discussion**

4 As in Studies 1 and 2, results were not consistent with hypotheses 1, 2 and 4:
5 consensus, authoritativeness, and new media literacy were not associated with self-
6 rated probability of liking the disinformation stories. In contrast to Studies 1 and 2,
7 however, conservatism did not predict liking the stories. Belief that the stories were
8 true was again the strongest predictor, while likelihood of having seen them before
9 was again statistically significant. Among the personality variables, lower
10 Agreeableness returned as a predictor of likely engagement with the stories,
11 consistent with Study 1 but not Study 2. Lower age predicted likely engagement, a
12 new finding, while being male predicted likely engagement as found in both in Study
13 1 and Study 2. Unlike Study 1 and Study 2, education had no effect.

14 With regard to historical accidental sharing, as in Study 3 higher Extraversion
15 was a predictor, while as in Study 1 so was lower Conscientiousness. Men were
16 more likely to have shared accidentally. Deliberate historical sharing was predicted
17 by higher levels of New Media Literacy. This is counter-intuitive and undermines the
18 argument that people share things because they know no better. In fact, in the
19 context of deliberate deception, motivated individuals higher in digital literacy may
20 actually be better equipped to spread untruths. Conservatism was also a predictor
21 here. This could again be a reflection of the consistency hypothesis, given that there
22 are high levels of conservative-oriented disinformation circulating. Finally, as in
23 Study 3, higher Neuroticism predicted deliberate historical sharing.

1 **Study 4**

2 Study 4 set out to repeat Study 1, but with a US sample and using US-centric
3 materials. The purpose of this was to test whether the observed effects applied
4 across different countries. The study was completed online, using as participants
5 members of research panels sourced through the research company Qualtrics.

6 **Method**

7 The methodology exactly replicated that of Study 1, except in the case of details
8 noted below. The planned analysis was revised to include the expanded set of
9 predictors eventually used in Study 1 (see Table 4).

10 **Materials**

11 Measures and materials were the same as used in Study 1. The only difference
12 from Study 1 was in the contents of the three disinformation exemplars, which were
13 designed to be relevant to a US rather than UK audience. Two of the stimuli were
14 sourced from the website Infowars.com, while a third was a story described as
15 untrue by the fact-checking website Politifact.com. In the same way as in Study 1,
16 the right-wing focus of the stories was again established in pilot work where a US
17 sample ($N=40$) saw seven stories including these and rated their political orientation
18 and likelihood of being shared. All were rated above the mid-point of an 11-point
19 scale asking “To what extent do you think this post was designed to appeal to people
20 with right wing (politically conservative) views?” anchored at “Very left wing oriented”
21 and “Very right wing oriented”. For the least right-wing of the three stories selected, a
22 one-sample t -test comparing the mean rating with the midpoint of the scale showed it
23 was statistically significantly higher, $t_{(39)}=6.729$, $p<.001$, $d=1.07$). One of the stimuli,
24 also used in Study 1-3, was titled “Revealed: UN plan to flood America with 600

1 million migrants". One was titled "Flashback: Obama's attack on internet freedom",
2 subtitled 'Globalists, Deep State continually targeting America's internet dominance',
3 featuring further anti- Obama, China and 'Big Tech' sentiment, and an image of
4 Barack Obama apparently drinking wine with a person of East Asian appearance.
5 The third was text based and featured material titled "Surgeon who exposed Clinton
6 foundation corruption in Haiti found dead in apartment with stab wound to the chest".

7 The materials used to manipulate authoritativeness (Facebook usernames
8 shown as sources of the stories) were the same as used in Studies 1-3. These were
9 retained because pilot work indicated that the higher and lower sets differed in
10 authoritativeness for US audiences in the same way as for UK audiences. A sample
11 of 30 US participants again each rated a selection of 9 usernames, including these 6,
12 for the extent to which each was "likely to be an authoritative source - that is, likely to
13 be a credible and reliable source of information". A within-subjects *t*-test indicated
14 that mean authoritativeness ratings for the 'higher' group were statistically
15 significantly higher than the 'lower' group ($t_{(29)}=-9.355$ $p<.001$, $d_z=1.70$).

16 **Procedure**

17 The procedure replicated Study 1, save that in this case the NMLS was
18 presented across two pages.

19 **Data screening and processing**

20 Prior to delivery of the sample, Qualtrics performed a series of quality checks
21 and 'data scrubbing' procedures to remove and replace participants with response
22 patterns suggesting inauthentic or inattentive responding. These included speeding
23 checks and examination of response patterns. On delivery of the initial sample
24 ($N=660$) further screening procedures were performed. Nine respondents were
25 identified who had responded with the same scores to substantive sections of the

1 questionnaire ('straightlining'), and one who had not completed any of the personality
2 items. Twelve respondents were not active Facebook users: Six reported using
3 Facebook 'not at all' and a further six less often than 'every few weeks'. All of these
4 were removed, leaving $N=638$. These checks and exclusions were carried out prior
5 to any data analysis.

6 **Participants**

7 The target sample size was planned to exceed $N=614$, as in Study 1. Qualtrics
8 was contracted to provide a sample of active Facebook users that was broadly
9 representative of the US population in terms of gender; education level; and age
10 profile (18+). Sampling quotas were used to assemble a sample comprising
11 approximately one third each self-describing as left-wing, centre and right-wing in
12 their political orientation. Sampling errors on the part of Qualtrics led to over-
13 recruitment of individuals aged 65 years, who make up 94 of the 160 individuals in
14 the 60-69 age group. As a consequence, the 60-69 age group is itself over-
15 represented in this sample compared to the broader US population. Participant
16 demographics are shown in Table 1, column 4. For the focal analysis in this study,
17 the sample size conferred 92.6% power to detect $R^2 = .04$ in a multiple regression
18 with 15 predictors (2-tailed, $\alpha = .05$).

19 **Results**

20 Descriptive statistics are summarised in Table 15. All scales had acceptable
21 reliability. The main dependent variable, likelihood of sharing, again had a very
22 skewed distribution with a strong floor effect.

23 **Table 15. Descriptive Statistics for Participant Characteristics and Response**
24 **Variables, Study 4.**

	<i>N</i>	<i>M</i>	<i>SD</i>	α	Range		Skew	Kurtosis
					Potential	Actual		
Openness	637	23.40	5.76	.74	7-35	9-35	0.05	-0.57
Conscientiousness	636	37.50	7.07	.81	10-50	15-50	-0.16	-0.67
Extraversion	635	27.51	6.84	.80	9-45	9-45	-0.15	0.14
Agreeableness	636	27.43	4.63	.71	7-35	10-35	-0.62	-0.24
Neuroticism	636	20.15	6.89	.82	8-40	8-40	0.30	-0.34
Conservatism	637	784.14	203.46	.83	0-1200	0-1200	-0.58	0.17
New Media Literacy	631	130.88	22.04	.95	35-175	40-171	-0.41	0.21
Likelihood of sharing	638	15.29	9.84		3-33	3-33	0.17	-1.25
Belief stories true	638	8.54	3.61		3-15	2-15	0.02	-1.00
Likelihood seen before	638	7.18	3.83		3-15	0-15	0.45	-1.02
Age	637	44.97	15.70			18-78	-0.02	-1.32

1

2

3

4

5

6

7

8

9

10

11

12

To simultaneously test hypotheses 1-4 a multiple regression analysis was carried out using the expanded predictor set from Study 1. Given inclusion of gender as a predictor variable, the one respondent who did not report their gender as either male or female was excluded from further analysis. The analysis, summarised in Table 16, indicated that the model explained 56% of the variance in self-reported likelihood of sharing the three disinformation items. Neither the authoritativeness of the story source, consensus information associated with the stories, nor consistency of the items with participant attitudes (conservatism) was a statistically significant predictor. Extraversion positively predicted ratings of likelihood of sharing. In terms of demographic characteristics, age was a significant predictor, with younger people

1 reporting a higher likelihood of sharing. Finally, people reported a greater likelihood
 2 of sharing the items if they believed they were likely to be true, and if they thought
 3 they had seen them before.

4 **Table 16. Predictors of Rated Likelihood of Sharing Stimuli (Extended**
 5 **Predictor Set), Study 4.**

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Constant	-7.335	3.923		-1.87	.062
New Media Literacy	0.015	0.014	0.033	1.014	.311
Conservatism	0.001	0.002	0.03	0.837	.403
Openness	-0.009	0.062	-0.006	-0.152	.879
Neuroticism	0.061	0.054	0.043	1.135	.257
Extraversion	0.12	0.046	0.084	2.607	.009
Conscientiousness	0.069	0.056	0.05	1.222	.222
Agreeableness	-0.12	0.08	-0.057	-1.489	.137
Authoritativeness	-0.337	0.532	-0.017	-0.633	.527
Consensus	-0.148	0.536	-0.008	-0.276	.783
Belief stories true	1.329	0.101	0.487	13.163	.000
Age	-0.053	0.021	-0.084	-2.577	.010
Gender (M=1, F=2)	-0.689	0.574	-0.035	-1.2	.231
Education	0.243	0.229	0.031	1.057	.291
Facebook use	0.694	0.36	0.055	1.93	.054
Likelihood seen before	0.605	0.094	0.234	6.455	.000
<i>R</i> ²			.560		
Adjusted <i>R</i> ²			.549		
<i>F</i>			51.577		.000

1 Authoritativeness and Consensus conditions coded as dummy variables;
2 higher=1, lower=0 in each case. $N=624$.

3

4 Participants had also been asked about their historical sharing of untrue political
5 stories, both unknowing and deliberate. Of the 638 participants, 185 (29.0%)
6 indicated that they had ever 'shared a political news story online that they later found
7 out was made up', while 132 out of 638 indicated they had shared one that they
8 'thought AT THE TIME was made up' (20.7%). Predictors of whether or not people
9 had shared untrue material under both sets of circumstances were examined using
10 logistic regressions, with the same sets of participant-level predictors.

11 Having unknowingly shared untrue material (Table 17) was significantly
12 predicted by higher New Media Literacy, lower Conscientiousness, higher education,
13 and higher levels of Facebook use. Having shared material known to be untrue at
14 the time (Table 18) was significantly predicted by higher Extraversion, lower
15 Agreeableness, younger age, and higher levels of Facebook use.

16 **Table 17. Binary Logistic Regression: Predictors of Whether Participants had**
17 **Previously Shared Political Stories They Subsequently Discovered Were False,**
18 **Study 4.**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp(B)</i>	<i>95% C.I. for EXP(B)</i>
New Media Literacy	0.015	0.005	7.936	1	.005	1.015	[1.005, 1.026]
Conservatism	0.000	0.001	0.466	1	.495	1.000	[0.999, 1.002]
Openness	0.012	0.022	0.275	1	.600	1.012	[0.969, 1.056]
Neuroticism	-0.005	0.020	0.072	1	.789	0.995	[0.956, 1.034]
Extraversion	0.020	0.017	1.366	1	.243	1.020	[0.987, 1.054]
Conscientiousness	-0.067	0.021	10.454	1	.001	0.935	[0.898, 0.974]

Agreeableness	-0.048	0.028	2.923	1	.087	0.953	[0.901, 1.007]
Age	-0.008	0.007	1.347	1	.246	0.992	[0.978, 1.006]
Gender (M=1, F=2)	0.068	0.207	0.108	1	.743	1.070	[0.713, 1.606]
Education	0.221	0.080	7.594	1	.006	1.247	[1.066, 1.459]
Facebook use	0.753	0.173	18.927	1	.000	2.124	[1.513, 2.982]
Constant	-4.115	1.495	7.577	1	.006	0.016	

1 *N*=624.

2

3 **Table 18. Binary Logistic Regression: Predictors of Whether Participants had**
4 **Previously Shared Political Stories They Knew at the Time Were False, Study**
5 **4.**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp(B)</i>	<i>95% C.I. for EXP(B)</i>
New Media Literacy	0.011	0.006	3.44	1	.064	1.011	[0.999, 1.023]
Conservatism	0.001	0.001	2.366	1	.124	1.001	[1.000, 1.003]
Openness	-0.024	0.026	0.855	1	.355	0.977	[0.929, 1.027]
Neuroticism	-0.008	0.024	0.102	1	.750	0.992	[0.947, 1.040]
Extraversion	0.045	0.020	4.88	1	.027	1.046	[1.005, 1.089]
Conscientiousness	-0.027	0.025	1.189	1	.276	0.973	[0.927, 1.022]
Agreeableness	-0.106	0.033	10.182	1	.001	0.900	[0.843, 0.960]
Age	-0.039	0.009	18.989	1	.000	0.962	[0.945, 0.979]
Gender (M=1, F=2)	-0.432	0.236	3.345	1	.067	0.649	[0.409, 1.031]
Education	0.148	0.088	2.853	1	.091	1.160	[0.976, 1.378]
Facebook use	0.344	0.173	3.972	1	.046	1.411	[1.006, 1.979]
Constant	-0.259	1.634	0.025	1	.874	0.772	

6 *N*=624.

1 **Discussion**

2 Again, the pattern of results emerging from Study 4 had some similarities but
3 also some differences from Studies 1-3. Once again, hypotheses 1, 2 and 4 were
4 unsupported by the results. Similarly to Study 3, but unlike Studies 1 and 2,
5 conservatism (the proxy for consistency) did not predict sharing the stories. Belief
6 that the stories were true, and likelihood of having seen them before, were the
7 strongest predictors. Higher levels of Extraversion (a new finding) and lower ages
8 (as in Study 3) were associated with higher reported likelihood of sharing the stimuli.

9 For historical sharing, for the first time – and counterintuitively – new media
10 literacy was associated with higher likelihood of having shared false material
11 unknowingly. As in Studies 1 and 2, lower Conscientiousness was also important.
12 Counterintuitively, higher education levels were associated with higher unintentional
13 sharing, as were higher levels of Facebook use. For intentional sharing, higher
14 Extraversion was a predictor, as was lower Agreeableness, younger age and higher
15 levels of Facebook use.

16 **General Discussion**

17 When interpreting the overall pattern of results from Studies 1-4, given the
18 weakness of most of the associations, it is likely to be most useful to focus on
19 relationships that are replicated across studies and disregard ‘one off’ findings.
20 Tables 19-21 provide a summary of the statistically significant predictors in each of
21 the studies. It is clear that two variables consistently predicted self-rated likelihood of
22 sharing disinformation exemplars: belief that the stories were likely to be true, and
23 likely prior familiarity with the stories. It is also clear that three key variables did not:
24 markers of authority, markers of consensus and digital literacy.

1 **Table 19. Predictors of Self-reported Likelihood of Sharing, Retweeting or**
 2 **Liking Stimuli Across all Studies.**

	Study 1 (Facebook sharing)	Study 2 (Twitter retweeting)	Study 3 (Instagram liking)	Study 4 (Facebook sharing)
New Media				
Literacy				
Conservatism	+	+		
Openness				
Neuroticism				
Extraversion				+
Conscientiousness				
Agreeableness	-		-	
Authoritativeness				
Consensus				
Belief stories true	+	+	+	+
Age			-	-
Gender (M=1, F=2)	-	-	-	
Education	-	-		
Platform use				

Likelihood seen before	+	+	+	+
------------------------	---	---	---	---

1 + denotes a statistically significant positive association, - a negative
2 association.

3

4 **Table 20. Predictors of Un-knowingly Sharing False Political Stories Across all**
5 **Studies.**

	Study 1 (Facebook sharing)	Study 2 (Twitter retweeting)	Study 3 (Instagram liking)	Study 4 (Facebook sharing)
New Media				+
Literacy				
Conservatism				
Openness				
Neuroticism				
Extraversion		+	+	
Conscientiousness	-		-	-
Agreeableness	-			
Age	-			
Gender (M=1, F=2)			-	
Education				+

Platform use + +

1 + denotes a statistically significant positive association, - a negative
 2 association.

3

4 **Table 21. Predictors of Knowingly Sharing False Political Stories Across all**
 5 **Studies.**

	Study 1 (Facebook sharing)	Study 2 (Twitter retweeting)	Study 3 (Instagram liking)	Study 4 (Facebook sharing)
New Media			+	
Literacy				
Conservatism			+	
Openness				
Neuroticism		+	+	
Extraversion				+
Conscientiousness				
Agreeableness	-			-
Age	-			-
Gender (M=1, F=2)		-		
Education				
Platform use				+

1 + denotes a statistically significant positive association, - a negative association.

2

3 Hypothesis 1 predicted that stories portrayed as coming from more authoritative
4 sources were more likely to be shared. However, this was not observed in any of the
5 four studies. One interpretation of this is that the manipulation failed. However, pilot
6 work (see Study 1, Study 4) with comparable samples indicated that people did see
7 the sources as differing in authoritativeness. The failure to find the predicted effect
8 could also be due to use of simulated scenarios – though care was taken to ensure
9 they resembled reality – or weaknesses in the methodology, such as the
10 distributional properties of the dependent variables. However, consistent
11 relationships between other predictors and the dependent variable were observed.
12 Thus, the current studies provide no evidence that authoritativeness of a source
13 influences sharing behaviour.

14 Hypothesis 2 predicted that stories portrayed as having a higher degree of
15 consensus in audience reactions (i.e. high numbers of people had previously shared
16 them) would be more likely to be shared. In fact, consensus markers had no effect
17 on self-reported probability of sharing or liking the stories. Therefore, the current
18 studies provide no evidence that indicators of ‘social proof’ influence participant
19 reactions to the stimuli.

20 Hypothesis 3 was that people would be more likely to share materials consistent
21 with their pre-existing beliefs. This was operationalised by measuring participants’
22 political orientation (overall level of conservatism) and using stimuli that were right-
23 wing in their orientation. In Studies 1 and 2, more conservative people were more
24 likely to share the materials. Further evidence for hypothesis 3 comes from the
25 finding, across all studies, that level of belief the stories were “accurate and truthful”

1 was the strongest predictor of likelihood of sharing. This is again in line with the
2 consistency hypothesis: people are behaving in ways consistent with their beliefs.
3 The finding from Study 3 that more conservative people were more likely to have
4 historically shared material they knew to be untrue could also be in line with this
5 hypothesis, given that a great many of the untrue political stories circulated online
6 are conservative-oriented.

7 Hypothesis 4, that people lower in digital literacy would be more likely to engage
8 with disinformation, was again not supported. As noted earlier, measurement of
9 digital literacy is problematic. However, pilot work showed that the New Media
10 Literacy Scale did differentiate between people with higher and lower levels of social
11 media use in the expected manner, so it is likely to have a degree of validity. In
12 Study 4, higher NMLS scores were associated with having unwittingly shared false
13 material in the past, which is counterintuitive. However, this may be due to the fact
14 that more digitally literate people should be more able to see that something was
15 false in hindsight. Higher NMLS scores were also associated with deliberately
16 sharing falsehoods in Study 3. This could be attributable to greater ease with which
17 digitally literate individuals can do such things, if motivated to do so.

18 A number of other variables were included on an exploratory basis, or for the
19 purpose of controlling for possible confounds. Of these, the most important was
20 participants' ratings of the likelihood that they had seen the stimuli before. This
21 variable was originally included in the design so that any familiarity effects could be
22 controlled for when evaluating the effect of other variables. In fact, rated likelihood of
23 having seen the materials before was the second strongest predictor of likelihood of
24 sharing it. It was a predictor in all four studies, and for the Facebook studies (1 and
25 4) it was the second most important variable. This is consistent with work on prior

1 exposure to false material online, where prior exposure to fake news headlines
2 increased participants' ratings of their accuracy [44]. Furthermore, it has been found
3 that prior exposure to fake-news headlines reduced participants' ratings of how
4 unethical it was to share or publish the material, even when it was clearly marked as
5 false [43]. Thus, repeated exposure to false material may increase our likelihood of
6 sharing it. It is known that repeated exposure to statements increases people's
7 subjective ratings of their truth [48]. However, there must be more going on here,
8 because the regression analyses indicated that the familiarity effect was independent
9 of the level of belief that it is true. When considering work that found that
10 amplification of content by bot networks led to greater levels of human sharing [21],
11 the implication is that repeated actual exposure to the materials is what prompts
12 people to share it, not metrics of consensus such as the number of likes or shares
13 displayed beside an article.

14 Of the five dimensions of personality measured, four (Agreeableness,
15 Extraversion, Neuroticism and Conscientiousness) were predictors of either current or
16 historical sharing in one or more studies. Consistent with findings from [24], Studies
17 1 and 3 found that lower Agreeableness was associated with greater probability of
18 sharing or liking the stories. It was also associated with accidental historical sharing
19 in Study 1, and deliberate historical sharing in Studies 1 and 4. In contrast to this,
20 past research on personality and social media behaviour indicates that *more*
21 agreeable people are more likely to share information on social media: [49] reported
22 that its role in this was mediated by trust, while [32] found that higher Agreeableness
23 was associated with higher levels of social media use in general. Given those
24 findings, it is likely that the current results are specific to disinformation stimuli rather
25 than social sharing in general. Agreeableness could potentially interact with the

1 source of the information: more agreeable people might conceivably be more eager
2 to please those close to them, However, while it is possible that Agreeableness
3 interacted in some way with the framing of the material having been shared by 'a
4 friend' in Study 1, Study 3 had no such framing. More broadly, the nature of the
5 stories may be important: disinformation items are normally critical or hostile in their
6 nature. This may mean they are more likely to be shared by disagreeable people,
7 who themselves may be critical in their outlook and not concerned about offending
8 others. Furthermore, Agreeableness is associated with general trusting behaviour. It
9 may be that disagreeable people are therefore more likely to endorse conspiracist
10 material, or other items consistent with a lack of trust in politicians or other public
11 figures.

12 Lower Conscientiousness was associated with accidental historical sharing of
13 false political stories in Studies 1, 3 and 4. This is unsurprising, as less conscientious
14 people would be less likely to check the veracity of a story before sharing it. The lack
15 of an association with deliberate historical sharing reinforces this view.

16 Higher Extraversion was associated with probability of sharing in Study 4, with
17 accidental historical sharing in Study 2 and 3, and with deliberate historical sharing in
18 Study 4. Higher Neuroticism was associated with historical deliberate sharing in
19 Studies 2 and 3. All these relationships may simply reflect a higher tendency on the
20 part of extraverted and neurotic individuals to use social media more [32].

21 There are clearly some links between personality and sharing of disinformation.
22 However, the relationships are weak and inconsistent across studies. It is possible
23 that different traits affect different behaviours: for example low Conscientiousness is
24 associated with accidental but not deliberate sharing, while high Neuroticism is
25 associated with deliberate but not accidental sharing. Thus, links between some

1 personality traits and the spread of disinformation may be context- and motivation-
2 specific, rather than reflecting blanket associations. However, lower Agreeableness –
3 and to a lesser extent higher Extraversion – may predict an overall tendency to
4 spread this kind of material.

5 Demographic variables were also measured and included in the analyses.
6 Younger individuals rated themselves as more likely to engage with the
7 disinformation stimuli in Studies 3 and 4, and were more likely to have shared untrue
8 political stories in the past either accidentally (Study 1) or deliberately (Studies 1 and
9 4). This runs counter to findings that older adults were much more likely to have
10 spread material from ‘fake news’ domains [16]. It is possible that the current findings
11 simply reflect a tendency of younger people to be more active on social media.

12 People with lower levels of education reported a greater likelihood of sharing the
13 disinformation stories in Studies 1 and 2. Counterintuitively, more educated people
14 were more likely to have accidentally shared false material in the past (Study 4). One
15 possible explanation is that more educated people are more likely to have realised
16 that they had done this, so the effect in Study 4 reflects an influence on reporting of
17 the behaviour rather than on the behaviour itself.

18 In each of Studies 1, 2 and 3, men reported a greater likelihood of sharing or
19 liking the stimuli. Men were also more likely to have shared false material in the past
20 unintentionally (Study 3) or deliberately (Study 2). Given its replicability, this would
21 seem to be a genuine relationship, but one which is not easy to explain.

22 Finally, the level of use of particular platforms (Facebook, Twitter or Instagram)
23 did not predict likelihood of sharing the stimuli in any study. Level of use of Twitter
24 (Study 2) predicted accidental sharing of falsehoods, while Facebook use predicted
25 both accidental and deliberate sharing (Study 4). For historical sharing, this may be

1 attributable to a volume effect: the more you use the platforms, the more likely you
2 are to do these things. It should be noted that the level of use metric lacked
3 granularity and had a strong ceiling effect, with most people reporting the highest
4 use level in each case.

5 In all four studies, a minority of respondents indicated that they had previously
6 shared political disinformation they had encountered online, either by mistake or
7 deliberately. The proportion who had done each varied across the four studies, likely
8 as a function of the population sampled (13.1%-29.0% accidentally; 6.2%-20.7%
9 deliberately), but the figures are a similar magnitude to those reported elsewhere
10 [31,16]. Even if the proportion of social media users who deliberately share false
11 information is just 6.2%, the lowest figure found here, then that is still a very large
12 number of people who are actively and knowingly spreading untruths.

13 The current results indicate that a number of variables predict onward sharing of
14 disinformation. However, most of these relationships are very small. It has been
15 argued that the minimum effect size for a predictor that would have real-world
16 importance in social science data is $\beta = .2$ [42]. Considering the effect sizes for the
17 predictors in Tables 4, 8, 12, and 17, only belief that the stories are true exceeds this
18 benchmark in every study, while probability of having seen the stories before
19 exceeded it in Studies 1 and 4. None of the other relationships reported exceeded
20 the threshold. This has implications for the practical importance of these findings, in
21 terms of informing interventions to counteract disinformation.

22 **Practical Implications**

23 Some of the key conclusions in this set of studies arise from the failure to find
24 evidence supporting an effect. Proceeding from such findings to a firm conclusion is
25 a logically dangerous endeavour: absence of evidence is not, of course, evidence of

1 absence. However, given the evidence from pilot studies that the manipulations were
2 appropriate; the associations of the dependent measures with other variables; and
3 the high levels of power to detect the specified effects, it is possible to say with some
4 confidence that hypotheses 1, 2 and 4 are not supported by the current data. This
5 means that the current project does not provide any evidence that interventions
6 based on these would be of value.

7 This is particularly important for the findings around digital literacy. Raising digital
8 media literacy is a common and appealing policy position for bodies concerned with
9 disinformation (e.g. [1]). There is evidence from a number of trials that it can be
10 effective in the populations studied. However, no support was found here for the idea
11 that digital literacy has a role to play in the spread of disinformation. This could
12 potentially be attributed to the methodology in this study. However, some participants
13 – 288 in total across all four studies – reported sharing false political stories that they
14 knew at the time were made up. It is hard to see how raising digital literacy would
15 reduce such deliberate deception. Trying to raise digital literacy across the
16 population is therefore unlikely to ever be a complete solution.

17 There is evidence that consistency with pre-existing beliefs can be an important
18 factor, especially in relation to beliefs that disinformation stories are accurate and
19 truthful. This implies that interventions are likely to be most effective when targeted
20 at individuals who already hold an opinion or belief, rather than trying to change
21 people's minds. While this would be more useful to those seeking to spread
22 disinformation, it could also give insights into populations worth targeting with
23 countermessages. Targeting on other variables – personality or demographic – is
24 unlikely to be of value given the low effect sizes. While these variables (perhaps

1 gender and Agreeableness in particular) most likely do play a role, their relative
2 importance seems so low that the information is unlikely to be useful in practice.

3 Alongside other recent work [43,44], the current findings suggest that repeated
4 exposure to disinformation materials may increase our likelihood of sharing it, even if
5 we don't believe it. The practical implication would be that to get a message
6 repeated online, one should repeat it many times (there is a clear parallel with the
7 'repeat the lie often enough' maxim regarding propaganda). Social proof (markers of
8 consensus) seems unimportant based on current findings, so there is no point in
9 trying to manipulate the numbers next to a post as sometimes done in online
10 marketing. What might be more effective is to have the message posted many times
11 (e.g. by bots) so that people had a greater chance of coming across it repeatedly.
12 This would be true both for disinformation and counter-messages.

13 **Limitations**

14 As a scenario-based study, the current work has a number of limitations. While it
15 is ethically preferable to field experiments, it suffers from reduced ecological validity
16 and reliance on self-reports rather than genuine behaviour. Questions could be
17 asked, for example, about whether the authoritativeness and consensus
18 manipulations were sufficiently salient to participants (even though they closely
19 mirrored the presentation of this information in real-life settings). Beyond this,
20 questions might be raised about the use of self-reported likelihood of sharing: does
21 sharing intention reflect real sharing behaviour? In fact, there is evidence to suggest
22 that it does, with recent work finding that self-reported willingness to share news
23 headlines on social media paralleled the actual level of sharing of those materials on
24 Twitter [50].

1 The scenarios presented were all selected to be right-wing in their orientation,
2 whereas participants spanned the full range from left to right in their political
3 attitudes. This means that consistency was only evaluated with respect to one pole
4 of the right-left dimension. There are a number of other dimensions that have been
5 used as wedge issues in real-world information operations: for example, support for
6 the Black Lives Matter movement; climate change; or for or against Britain leaving
7 the European Union. The current research only evaluated consistency between
8 attitudes and a single issue. A better test of the consistency hypothesis would be to
9 extend that to evaluation of consistency between attitudes and some of those other
10 issues.

11 A key issue is the distributions of the main outcome variables, which were
12 heavily skewed with strong floor effects. While they still had sufficient sensitivity to
13 make the regression analyses meaningful, they also meant that any effects found
14 were likely to be attenuated. It may thus be that the current findings underestimate
15 the strength of some of the associations reported.

16 Another measurement issue is around the index of social media use (Facebook,
17 Twitter, Instagram). As Table 1 shows, in three of the studies over 60% of
18 respondents fall into the highest use category. Again, this weakens the sensitivity of
19 evaluations of these variables as predictors of sharing disinformation.

20 In order to identify variables associated with sharing disinformation, this research
21 programme took the approach of presenting individuals with examples of
22 disinformation, then testing which of the measured variables was associated with
23 self-reported likelihood of sharing. A shortcoming of this approach is that it does not
24 permit us to evaluate whether the same variables are associated with sharing true
25 information. An alternative design would be to show participants either true or false

1 information, and examine whether the same constructs predict sharing both. This
2 would enable identification of variables differentially impacting the sharing of
3 disinformation but not true information. Complexity arises, however, from the fact
4 that whether a story can be considered disinformation, misinformation, or true
5 information, depends on the observer's perspective. False material deliberately
6 placed online would be categorized as disinformation. A social media user sharing it
7 in full knowledge that it was untrue would be sharing disinformation. However, if they
8 shared it believing it was actually true, then from an observer's perspective this
9 would be technically categorised as misinformation (defined as "the inadvertent
10 sharing of false information" [1, p.10]). In fact, from the user's perspective, it would
11 be true information (because they believe it) even though an omniscient observer
12 would know it was actually false. This points to the importance of further research
13 into user motivations for sharing, which are likely to differ depending on whether or
14 not they believe the material is true.

15 In three of the four studies (Studies 1,2, 4), the stimulus material was introduced
16 as having been posted by a friend who wanted them to share it. This is likely to have
17 boosted the rates of self-reported likelihood of sharing in those studies. Previous
18 work has shown that people rate themselves as more likely to engage with potential
19 disinformation stories posted by a friend, as opposed to a more distant acquaintance
20 [24]. To be clear, this does not compromise the testing of hypotheses in those
21 studies (given that the framing was the same for all participants, in all conditions). It
22 is also a realistic representation of how we may encounter material like this in our
23 social media feeds. However, it does introduce an additional difference between
24 Studies 1, 2 and 4 when compared with Study 3. It would be desirable for further

1 work to check whether the same effects were found when messages were framed as
2 having been posted by people other than friends.

3 Finally, the time spent reading and reacting to the disinformation stimuli was not
4 measured. It is possible that faster response times would be indicative of more use
5 of heuristics rather than considered thought about the issues. This could profitably
6 be examined, potentially in observational or simulation studies rather than using self-
7 report methodology.

8 **Future Work**

9 A number of priorities for future research arise from the current work. First, it is
10 desirable to confirm these findings using real-world behavioural measures rather
11 than simulations. While it is not ethically acceptable to run experimental studies
12 posting false information on social media, it would be possible to do real-world
13 observational work. For example, one could measure digital literacy in a sample of
14 respondents, then do analyses of their past social media sharing behaviour.

15 Another priority revolves around those individuals who knowingly share false
16 information. Why do they do this? Without understanding the motivations of this
17 group, any interventions aimed at reducing the behaviour are unlikely to be
18 successful. As well as being of academic interest, motivation for sharing false
19 material has been flagged as a gap in our knowledge by key stakeholders [7].

20 The current work found that men were more likely to spread disinformation than
21 women. At present, it is not clear why this was the case. Are there gender-linked
22 individual differences that influence the behaviour? Could it be that the subject
23 matter of disinformation stories is stereotypically more interesting to men, or that
24 men think their social networks are more likely to be interested in or sympathetic to
25 them?

1 While the focus in this paper has been on factors influencing the spread of
2 untruths, it should be remembered that ‘fake news’ is only one element in online
3 information operations. Other tactics and phenomena, such as selective or out-of-
4 context presentation of true information, political memes, and deliberately polarising
5 hyperpartisan communication, are also prevalent. Work is required to establish
6 whether the findings of this project related to disinformation, also apply to those other
7 forms of computational propaganda. Related to this, it would be of value to establish
8 whether the factors found here to influence sharing of untrue information, also
9 influence the sharing of true information. This would indicate whether there is
10 anything different about disinformation, and also point to factors that might influence
11 sharing of true information that is selectively presented in information operations.

12 **Conclusion**

13 The current work allows some conclusions to be drawn about the kind of people
14 who are likely to further spread disinformation material they encounter on social
15 media. Typically, these will be people who think the material is likely to be true, or
16 have beliefs consistent with it. They are likely to have previous familiarity with the
17 materials. They are likely to be younger, male, and less educated. With respect to
18 personality, it is possible that they will tend to be lower in Agreeableness and
19 Conscientiousness, and higher in Extraversion and Neuroticism. With the exception
20 of consistency and prior exposure, all of these effects are weak and may be
21 inconsistent across different populations, platforms, and behaviours (deliberate v.
22 innocuous sharing). The current findings do not suggest they are likely to be
23 influenced by the source of the material they encounter, or indicators of how many
24 other people have previously engaged with it. No evidence was found that level of
25 literacy regarding new digital media makes much difference to their behaviour.

1 These findings have implications for how governments and other bodies should go
2 about tackling the problem of disinformation in social media.

3

4 **References**

5 1. House of Commons Digital, Culture, Media and Sport Committee.

6 Disinformation and 'fake news': Final Report. 2019 Feb 2 [cited 18 Feb
7 2019]. Available from:

8 <https://publications.parliament.uk/pa/cm201719/cmselect/cmcomeds/1791/1791.pdf> .

10 2. Bradshaw S, Howard PN. The Global Disinformation Order: 2019 Global

11 Inventory of Organised Social Media Manipulation. 2019 [cited 26

12 September 2019]. Available from: [https://comprop.oii.ox.ac.uk/wp-](https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf)

13 [content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf](https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf) .

14 3. Krasodonski-Jones A, Judson E, Smith J, Miller C, Jones E. Warring Songs:

15 Information operations in the digital age. 2019 May [cited 21 May 2019].

16 Available from: [https://demos.co.uk/wp-content/uploads/2019/05/Warring-](https://demos.co.uk/wp-content/uploads/2019/05/Warring-Songs-final-1.pdf)

17 [Songs-final-1.pdf](https://demos.co.uk/wp-content/uploads/2019/05/Warring-Songs-final-1.pdf) .

18 4. Howard PN, Ganash B, Liotsiou D, Kell J, François C. The IRA, Social Media

19 and Political Polarization in the United States, 2012-2018. Working

20 Paper 2018.2. 2018 [cited 20 December 2019]. Available from

21 [https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report-](https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report-2018.pdf)

22 [2018.pdf](https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report-2018.pdf) .

23 5. Zerback T, Töpfl F, Knöpfle M. The disconcerting potential of online

24 disinformation: Persuasive effects of astroturfing comments and three

- 1 strategies for inoculation against them. *New Media & Society*. 2020. Available
2 from: <https://doi.org/10.1177/1461444820908530> .
- 3 6. Parveen N, Waterson J. UK phone masts attacked amid 5G-coronavirus
4 conspiracy theory. *The Guardian*. 2020 April 4 [cited 2020 July 17]. Available
5 from: [https://www.theguardian.com/uk-news/2020/apr/04/uk-phone-masts-](https://www.theguardian.com/uk-news/2020/apr/04/uk-phone-masts-attacked-amid-5g-coronavirus-conspiracy-theory)
6 [attacked-amid-5g-coronavirus-conspiracy-theory](https://www.theguardian.com/uk-news/2020/apr/04/uk-phone-masts-attacked-amid-5g-coronavirus-conspiracy-theory) .
- 7 7. House of Commons Digital, Culture, Media and Sport Committee. Misinformation
8 in the COVID-19 Infodemic. 2020 July 21 [cited 21 July 2020]. Available from:
9 <https://committees.parliament.uk/publications/1954/documents/19089/default/> .
- 10 8. Bradshaw S, Neudert L-M, Howard PN. Government responses to malicious
11 use of social media. 2018 [cited 17 January 2019]. Available from:
12 [https://www.stratcomcoe.org/government-responses-malicious-use-social-](https://www.stratcomcoe.org/government-responses-malicious-use-social-media)
13 [media](https://www.stratcomcoe.org/government-responses-malicious-use-social-media) .
- 14 9. Facebook. What's the difference between organic, paid and post reach?
15 2019 [cited 31 July 2019]. Available from:
16 <https://www.facebook.com/help/285625061456389> .
- 17 10. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*.
18 2018;359:1146-1151.
- 19 11. Petty RE, Cacioppo JT. The Elaboration Likelihood Model of Persuasion. In:
20 Berkowitz L, editor. *Advances in Experimental Social Psychology Volume 19*.
21 Academic Press; 1986. p. 123-205.
- 22 12. Hayes RA, Carr CT, Wohn DY. One Click, Many Meanings: Interpreting
23 Paralinguistic Digital Affordances in Social Media. *Journal of Broadcasting &*
24 *Electronic Media*. 2016;60:171-187.

- 1 13. Williams EJ, Beardmore A, Joinson AN. Individual differences in susceptibility
2 to online influence: A theoretical review. *Computers in Human Behavior*.
3 2017;72: 412-421.
- 4 14. Cook J, Lewandowsky S, Ecker UKH. Neutralizing misinformation through
5 inoculation: Exposing misleading argumentation techniques reduces their
6 influence. *PLoS One*. 2017;12: e0175799.
- 7 15. Cialdini RB. *Influence: The Psychology of Persuasion*. New York:
8 HarperCollins; 2009.
- 9 16. Guess A, Nagler J, Tucker J. Less than you think: Prevalence and predictors of
10 fake news dissemination on Facebook. *Science Advances*. 2019;5: eaau4586.
- 11 17. Festinger L. *A theory of cognitive dissonance*. Stanford University Press; 1957.
- 12 18. Moravec, PL, Minas, RK, Dennis AR. Fake News on Social Media: People
13 Believe What They Want to Believe When it Makes No Sense At All. *MIS*
14 *Quarterly*. 2019;43: 13430-11360.
- 15 19. Roethke K, Klumpe J, Adam M, Benlian A. Social influence tactics in e-
16 commerce onboarding: The role of social proof and reciprocity in affecting user
17 registrations. *Decision Support Systems*. 2020. doi:
18 10.1016/j.dss.2020.113268.
- 19 20. Innes M, Dobрева D, Innes H. Disinformation and digital influencing after
20 terrorism: spoofing, truthing and social proofing. *Contemporary Social Science*.
21 2019. doi: 10.1080/21582041.2019.1569714.
- 22 21. Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F. The
23 spread of low-credibility content by social bots. *Nature Communications*.
24 2018;9: 4787.

- 1 22. Bay S, Fredheim R. How Social Media Companies are Failing to Combat
2 Inauthentic Behaviour Online. 2019 [cited 21 February 2020]. Available from:
3 [https://stratcomcoe.org/how-social-media-companies-are-failing-combat-
5 inauthentic-behaviour-online](https://stratcomcoe.org/how-social-media-companies-are-failing-combat-
4 inauthentic-behaviour-online) .
- 5 23. Lin X, Spence PR, Lachlan KA. Social media and credibility indicators: The
6 effect of influence cues. *Computers in Human Behavior*. 2016;63: 264-271.
- 7 24. Buchanan T, Benson V. Spreading Disinformation on Facebook: Do Trust in
8 Message Source, Risk Propensity, or Personality Affect the Organic Reach of
9 “Fake News”. *Social Media + Society*. 2019;5: 1-9.
- 10 25. Mak T, Berry L. Russian Influence Campaign Sought To Exploit
11 Americans’ Trust In Local News. 2018 July 12 [cited 8 August 2018].
12 Available from: [https://www.npr.org/2018/07/12/628085238/russian-influence-
14 campaign-sought-to-exploit-americans-trust-in-local-news](https://www.npr.org/2018/07/12/628085238/russian-influence-
13 campaign-sought-to-exploit-americans-trust-in-local-news) .
- 14 26. Mitchell A, Gottfried J, Barthel M, Shearer E. The Modern News Consumer:
15 News attitudes and practices in the digital era. 2016 [cited 13 November
16 2018]. Available from: [http://www.journalism.org/2016/07/07/the-modern-news-
18 consumer/](http://www.journalism.org/2016/07/07/the-modern-news-
17 consumer/) .
- 18 27. Zuiderveen Borgesius FJ, Möller J, Kruijkemeier S et al. Online Political
19 Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*.
20 2018;14: 82.
- 21 28. Chetty K, Qigui L, Gcora N, Josie J, Wenwei L, Fang C. Bridging the digital
22 divide: measuring digital literacy. *Economics: The Open-Access, Open-
23 Assessment E-Journal*. 2018;12: 1-20.

- 1 29. Koc M, Barut E. Development and validation of New Media Literacy Scale
2 (NMLS) for university students. *Computers in Human Behavior*. 2016;63: 834-
3 843.
- 4 30. Vishwanath A, Herath T, Chen R, Wang J, Rao HR. Why do people get
5 phished? Testing individual differences in phishing vulnerability within an
6 integrated, information processing model. *Decision Support Systems*. 2011;51:
7 576-586.
- 8 31. Barthel M, Mitchell A, Holcomb J. Many Americans Believe Fake News is
9 Sowing Confusion. 2016 Dec [cited 15 March 2018]. Available from:
10 [http://assets.pewresearch.org/wp-](http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/12/14154753/PJ_2016.12.15_fake-news_FINAL.pdf)
11 [content/uploads/sites/13/2016/12/14154753/PJ_2016.12.15_fake-](http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/12/14154753/PJ_2016.12.15_fake-news_FINAL.pdf)
12 [news_FINAL.pdf](http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/12/14154753/PJ_2016.12.15_fake-news_FINAL.pdf) .
- 13 32. Gil de Zúñiga H, Diehl T, Huber B, Liu J. Personality Traits and Social Media
14 Use in 20 Countries: How Personality Relates to Frequency of Social Media
15 Use, Social Media News Use, and Social Media Use for Social Interaction.
16 *Cyberpsychology, Behavior, and Social Networking*. 2017;20: 540-552.
- 17 33. Azucar D, Marengo D, Settanni M. Predicting the Big 5 personality traits from
18 digital footprints on social media: A meta-analysis. *Personality and Individual*
19 *Differences*. 2018;124: 150-159.
- 20 34. Hinds J, Joinson A. Human and Computer Personality Prediction From Digital
21 Footprints. *Current Directions in Psychological Science*. 2019;28: 204-211.
- 22 35. Matz SC, Kosinski M, Nave G, Stillwell DJ. Psychological targeting as an
23 effective approach to digital mass persuasion. *Proceedings of the National*
24 *Academy of Sciences of the United States of America*. 2017;114: 12714-12719.

- 1 36. Hall JA, Pennington N, Lueders A. Impression management and formation on
2 Facebook: A lens model approach. *New Media & Society*. 2013;16: 958-982.
- 3 37. Goldberg LR. A broad-bandwidth, public domain, personality inventory
4 measuring the lower-level facets of several five-factor models. In: I. Mervielde,
5 I.J. Deary, F. De Fruyt, FO, editors. *Personality Psychology in Europe Vol. 7*.
6 Tilburg, The Netherlands: Tilburg University Press; 1999. p. 7-28.
- 7 38. Buchanan T, Johnson JA, Goldberg LR. Implementing a Five-Factor
8 Personality Inventory for Use on the Internet. *European Journal of*
9 *Psychological Assessment*. 2005;21: 115-127.
- 10 39. Costa PT, McCrae RR. Revised NEO Personality Inventory (NEO PI-R) and
11 NEO Five-Factor Inventory (NEO FFI): Professional Manual. Odessa, FL:
12 Psychological Assessment Resources; 1992.
- 13 40. Everett JA. The 12 item Social and Economic Conservatism Scale (SECS).
14 *PLoS One*. 2013;8: e82131.
- 15 41. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on
16 Twitter during the 2016 U.S. presidential election. *Science*. 2019;363: 374-378.
- 17 42. Ferguson CJ. An effect size primer: A guide for clinicians and researchers.
18 *Professional Psychology: Research and Practice*. 2009;40: 532-538.
- 19 43. Effron DA, Raj M. Misinformation and Morality: Encountering Fake-News
20 Headlines Makes Them Seem Less Unethical to Publish and Share.
21 *Psychological Science*. 2020;31: 75-87.
- 22 44. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived
23 accuracy of fake news. *Journal of Experimental Psychology: General*.
24 2018;147: 1865-1880.

- 1 45. Liberini F, Redoano M, Russo A, Cuevas A, Cuevas R. Politics in the Facebook
2 Era. Evidence from the 2016 US Presidential Elections. CAGE Working Paper
3 Series (389). 2018 [cited 17 Dec 2019]. Available from:
4 [https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/public](https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/389-2018_redoano.pdf)
5 [ations/389-2018_redoano.pdf](https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/389-2018_redoano.pdf) .
- 6 46. Meehl PE. Why summaries of research on psychological theories are often
7 uninterpretable. *Psychological Reports*. 1990;66: 195-244.
- 8 47. DiResta R, Shaffer K, Ruppel B et al. The tactics & tropes of the Internet
9 Research Agency. 2018 [cited 12 Dec 2018]. Available from:
10 [https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senat](https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs)
11 [edocs](https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs) .
- 12 48. Dechêne A, Stahl C, Hansen J, Wänke M. The truth about the truth: a meta-
13 analytic review of the truth effect. *Personality and Social Psychology Review*.
14 2020;14: 238-257.
- 15 49. Deng S, Lin Y, Liu Y, Chen X, Li H. How Do Personality Traits Shape
16 Information-Sharing Behaviour in Social Media? Exploring the Mediating Effect
17 of Generalized Trust. *Information Research: An International Electronic*
18 *Journal*. 2017;22. Available from: <http://informationr.net/ir/22-3/paper763.html> .
- 19 50. Mosleh M, Pennycook G, Rand DG. Self-reported willingness to share political
20 news articles in online surveys correlates with actual sharing on Twitter. *PloS*
21 *One*. 2020;15: e0228882.
- 22