Volume 21, N° 1,
Summer 2021

PolSci
Romanian Journal of Political Science

Romanian Journal of Political Science

Revista Română de Științe Politice

The Romanian Academic Society (RAS)

ISSN 1582-456X

Cover photo by rawpixel.com
Source: freepik.com

ROMANIAN JOURNAL OF POLITICAL SCIENCE

Revista Română de Științe Politice

Volume 21, Nº 1, Summer 2021

TABLE OF CONTENTS

POLSCI PAPERS

Is social media affecting the perceived trustworthiness of misinformation?

Evidence from experimental comparisons

Christoph M. Abels[38]

Abstract

Social media plays a major role in the dissemination of misinformation. Although social networking sites (SNS) largely differ in their functionality and appearances, and therefore their ability to serve as misinformation vector, researchers have rarely systematically compared different SNS to investigate platform-specific effects beyond Facebook and Twitter.

This study tries to address this lack of diversity concerning SNS in the literature by experimentally comparing SNS-specific effects across seven different platforms (Discord, Facebook, Instagram, LinkedIn, Telegram, Twitter, WhatsApp) and the Associated Press (AP) website as control condition. The focus is on the perceived trustworthiness of manipulated news, as well as on participants' willingness to share, interact and distribute this news to friends or family. While platforms' specific effects do not vary significantly in this experiment, further analysis shows that prior exposure to misinformation affects credibility judgements across all of them and could thereby inform an evidence-based strategy against SNS-powered misinformation.

Keywords: Social media, Misinformation, Fake news, Trustworthiness, WhatsApp, Telegram, LinkedIn, Instagram

---

[38] Christoph Abels is a PhD researcher at the Hertie School. He holds a Bachelor of Science in Psychology from the University of Hagen and a Master of Public Policy from the Hertie School. His professional experience ranges from positions at various Berlin start-ups to consultancy work, having worked at Rocket Internet, Ketchum Pleon and Hering Schuppener. He is currently employed as a Research Associate at the Hertie School and leads the digitalisation and cybersecurity research programme for Polis180, a Berlin-based think tank. In his doctoral thesis, he is investigating the phenomenon of fake news.

1. Introduction

Misinformation has emerged as a major challenge for societies around the globe – from false or misleading information surrounding elections and referenda (Lazer et al., 2018), to potentially harmful health claims (Swire-Thompson & Lazer, 2020), and falsehoods undermining scientific findings (Lewandowsky et al., 2013, 2015).

The proliferation of these claims – in many cases incorporated into a specific conspiratorial belief system– may therefore inform decision-making on an individual level.

These claims are frequently spread using Social Network Sites (SNS), understood in the sense of Boyd and Ellison (2007) as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (p. 211). Although SNS initially started as a straightforward means of staying in touch with friends, their appropriation as political communication tool now garners significant attention from scholars and the public. SNS differ in their functionality, user base, and their intended use.

They also differ in their past involvement in spreading misinformation. With its significant role as misinformation source during the 2016 US presidential elections (Guess et al., 2020), Facebook became known for its failure to proactively address misinformation within its platform.

With the emergence of the COVID-19 pandemic, Facebook has again shown itself to be a potent tool for spreading misinformation. False or misleading claims are widely shared in invite-only Facebook groups, encompassing issues such as claims on the mismanagement of public authorities in responding to the pandemic; stories which characterize refugees as 'patient zeros' bringing the virus to Europe; or 5G networks as being true cause of the symptoms caused by COVID-19 (Scott, 2020).

Facebook is hardly the only SNS struggling to control misinformation. By the end of March 2020, Twitter became the platform with the highest number of false social posts according to a study conducted by Brennen, Simon, Howard, and Nielsen (2020). 59% of these posts involve information that is "spun, twisted, recontextualised, or reworked" (p. 1), whereas another 38% were found to be entirely fabricated. In examining content shared via WhatsApp, Garimella and Eckles (2020) investigated images distributed on public WhatsApp groups in India. They found that 13% of these images can be considered misinformation, with images taken out of context, manipulated (e.g., with Photoshop), or being used as misleading memes (images with added text) that may alter the intended meaning of the original image.

With the rising popularity and an ever-increasing number of different SNS, misinformation agents have a growing arsenal at their disposal to systematically spread misinformation. When SNS struggle to contain misinformation, their last firewall is then the recipients of the misleading contents, who can avoid being hoodwinked by the falsehoods they encounter. That firewall, however, is holey at best, as humans can fall victim to different cognitive biases and constraints that make them susceptible to misinformation.

People as individuals are more likely to believe information that is congruent with their own pre-existing attitudes (Taber & Lodge, 2006; Taber & Young, 2013) and will frequently fail to think analytically. This thereby increases individual susceptibility to misinformation (Pennycook & Rand, 2019) and people will experience difficulties in directing their attention to relevant aspects of information, such as source and plausibility.

Some authors also evaluate personality traits and psychological dispositions that make people more susceptible to adopting conspiracy beliefs – irrespective of whether these beliefs are related or not. Trust in others and a belief in an external locus of control (that many events are beyond individual control) are both negatively associated with conspiracy beliefs (Abalakina-Paap et al., 1999). Imhoff and Bruder (2014) postulate a conspiracy mentality, leading people to be more sceptical about those in power, a tendency to blame these groups for negative events, and an increased likelihood to engage in behaviour that is intended to challenge the status quo. Conspiracy beliefs are also rarely held singularly – belief in one conspiracy is associated with adoption of further conspiracies (Swami et al., 2010). Furthermore, Meyer et al. (2021) provides evidence that traits which interfere with the acquisition, maintenance, and transmission of knowledge, i.e., prejudice and close-mindedness, increases the susceptibility to misinformation. These traits, called 'epistemic vices', are found to have a stronger association with misinformation susceptibility than "political identity, educational attainment, scores on the Cognitive Reflection Test, personality, dogmatism, and need for closure" (p. 1).

Even after exposure to misleading claims it may prove difficult to correct or debunk these, as people tend to not adjust their memory and include new information in their considerations (Johnson & Seifert, 1994; Lewandowsky et al., 2012).

Misinformation should be addressed structurally at the SNS-level before users are exposed to it. Trying to understand if and how contextual elements of different SNS affect their ability to serve as misinformation vectors is therefore a worthwhile building block in current efforts to develop evidence-based policy responses. Especially as there appears to be a lack of coverage when it comes to comparative / comprehensive SNS research: more than two-thirds of studies on SNS address only a single platform, with Facebook being that platform in roughly 80% of cases (Rains & Brunner, 2015).

This paper contributes to this literature by asking:

Do different SNS channels vary in influence over the perceived trustworthiness of news and the respondents' interactions with news (such as sharing, interacting, forwarding, etc.)?

I hypothesize that the more trustworthy a SNS is perceived to be, the more trustworthy individual news items that are posted to the SNS are deemed to be (that is, that platform and post trustworthiness are positively linearly corelated). Given the reputational scandals affecting several global SNS (e.g., Cambridge Analytica in relation to Facebook), I envisage that those SNS whose reputations are subject to public scrutiny experience a decline in trustworthiness of news items. Additionally, I expect that political knowledge and trust in traditional media have a strong influence on trustworthiness perceptions. Knowledge has been identified as an important factor for trustworthiness judgments in social media environments (Sterrett et al., 2019), while the degree of trust in traditional media jointly with the frequency of media use serves as proxy for individuals' willingness to engage with media.

This paper is then organized as follows: in the first section I outline the role of credibility cues for judging SNS credibility. In the second section, I introduce the methodology of the experimental survey and the main variables of analysis. Subsequently, in the third section, I present and discuss the main result, before concluding with some of the implications of my findings.

2. Theory

Credibility cues can be understood as properties of the digital environment.

Credibility can be understood as believability, based on the perceived trustworthiness as well as expertise of a source or message (Flanagin et al., 2020; Rieh & Danielson, 2007). Source can be understood according to Sundar and Nass (2001) as everything in the chain of sender/presenter, medium/channel, as well as receiver/audience serves as a source – and can therefore by assessed in terms of its credibility.

Much of the source credibility literature looks at visible sources because the psychological effects of who presents the content is considered more powerful than that of who publishes the content. Credibility concepts that investigate the properties of SNS, which serve as a channel presenting the news, therefore offer a starting point to investigate SNS credibility effects. Investigating digital media, Flanagin and Metzger (2007) identify three types of credibility: message credibility, site credibility and sponsor credibility. While message credibility relies on characteristics of the message itself (i.e., accuracy, currency, information quality), site credibility depends on a site's visual design, the density of information presented, as well as the interactivity offered by the website. If a website is sponsored its credibility might be also influenced by public perceptions about the sponsor, such as a sponsor's reputation or personal experience.

Beyond these properties, Tseng and Fogg (1999) differentiate four types of credibility that include aspects of design and social recommendation: presumed, reputed, surface, and experienced credibility. General assumptions about the origin of the information, e.g., that politicians cannot be trusted in general, fall under the label of presumed credibility. In contrast, reputed credibility does not concern own assumptions about the source, but "what third parties have reported" (p. 42). Academic titles granted by prestigious institutions, such as doctors or professors, tend to increase the individual credibility of its bearer. As Tseng and Fogg argue, reputed credibility is especially pervasive online as different SNS and websites cross-reference each other, which can be interpreted as a third-party endorsement. Surface credibility refers to the perceived credibility of a person or object based on inspection. As Tseng and Fogg describe it "people are judging a book by its cover" (p. 42). Experience credibility is based on past interactions with a person or object and thereby empirically informed. It is the most complex type among the four, as it is built over time in an iterative process.

Differing investments of cognitive resources when encountering information have been discussed under the umbrella of dual processing models, such as the Elaboration Likelihood Model, the Controlled versus Automatic Processing Models (Shiffrin & Schneider, 1977), Heuristic-Systematic Model (Chaiken, 1980) and the Two Systems model (Kahnemann, 2011).

These models share the idea that information is processed in two different ways: an attentive, slow and resource-demanding process, and a fast, efficient way that demands less cognitive effort. Information processing is therefore a function of both cognitive ability and motivation of the receiver (Metzger, 2007). In situations where less motivation is involved and cognitive ability is low, judgments about a message or source are more likely to be made based on heuristics. Reputation, for example, can serve as heuristic to judge credibility (Metzger et al., 2010). Therefore, superficial properties of source or message become more important in the decision-making process. In relation to Tseng and Fogg's (1999) conceptualization, reputed and surface credibility might be especially salient heuristics.

Several findings show that the design of SNS and the resulting affordances which shape the type of content and the way it is presented on a platform, as well as the users cognitive state when interacting with the SNS (Sundar, 2008), might enable this superficial processing. Pearson (2020) provides evidence that blending both news and entertainment content, as is common on many SNS, makes it more likely that all content is processed inattentively. He argues that this environment may also increase belief in misinformation. Source layering furthermore complicates information processing (Sundar & Nass, 2001), which is conceptualized by Kang et al. (2011) through the idea of psychological distance to a source. Given that many news outlets, i.e., *The New York Times*, *the Washington Post*, *Wall Street Journal*, etc, share their content on several SNS, these outlets can be seen as distal sources – and one layer in the source hierarchy. The SNS themselves on which the news is shared constitute another, more proximate layer to the reader – and more proximate source cues can have a greater influence on the message's credibility than distal sources.

Kang et al. propose that differences in influence results from news consumers belief that the proximate source (the SNS) might be the actual source (the news outlet) in cases where consumers do not invest the cognitive resources needed to distinguish proximate and distal sources. This is supported by earlier evidence showing that four identical news stories were perceived differently, depending on the source the news was attributed to (proximate or distal) (Sundar & Nass, 2001).

If differences in SNS properties impact the perception of the platform, SNS might also differ in their ability to be used as misinformation vector. A few studies which assessed these differences indeed provide some evidence for this hypothesis. Research by Vraga and Bode (2018), who tested the effectiveness of social correction on misinformation about the Zika virus, offers a picture of complex interactions between social media platform and correcting interventions: When a source is added to a Facebook comment, the evaluation of that comment on Facebook is increased (i.e., it is judged as being more credible, trustworthy, accurate, etc.).

This effect is not observed for Twitter. Beyond that, these perceptions do not translate into reduced misperceptions of the causes of the Zika virus on Facebook, while higher evaluations of Twitter replies are associated with reduced misperceptions. Beyond the structural elements, behavioural differences are also observed: respondents in a Brazilian survey reported that they experience, witness, and engage in social interactions more on WhatsApp than on Twitter, underlining behavioural differences when engaging among platforms (Rossini et al., 2020). These findings illustrate the complex pattern of structural and behavioural interacting factors that impact the effect of misinformation.

3. Method

Starting with the research design, the following section describes the methodology applied in this paper. Stimulus material, dependent variables and covariates, as well as recruitment approach and sample size are summarized. Lastly, the procedure is illustrated.

Research design

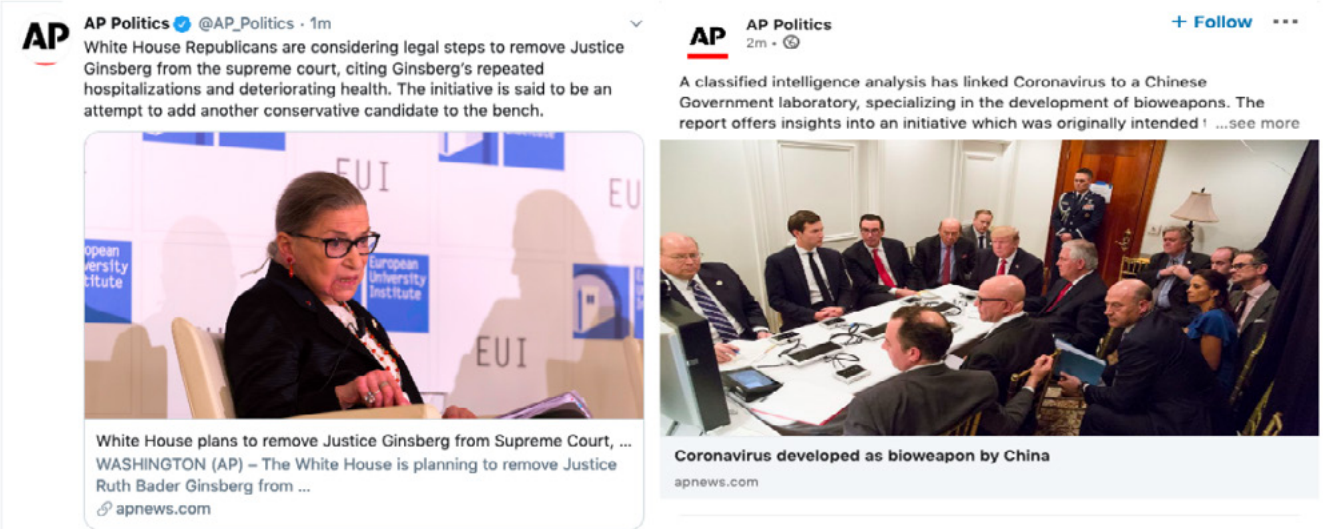
The study follows a between-subjects design, comparing seven different SNS as well as the AP website as control group (Discord, Facebook, Instagram, LinkedIn, Twitter, Telegram, WhatsApp, AP Website). Participants were randomly assigned to both factors.

Participants were told the following: “Below is a screenshot taken from [Discord, Facebook, Instagram, etc.]. Please rate its trustworthiness, and indicate whether you would like, share, and forward it. “

Stimulus material

Mock SNS designs were created to imitate the seven SNS and the AP website. The items had the appearance of a screenshot taken from the platform, immediately after it was posted, to avoid time effects. All relevant popularity indicators, such as the number of likes or shares, are set to zero. 14 items were pre-tested out of which 7 were included in the study. Figure 2 gives examples of the stimulus material. AP, which is shown on every item, was chosen to provide a credible and comparatively neutral mainstream news source.

Figure 2: Examples of stimulus material - Twitter (left) and LinkedIn (right)



Behavioural responses and covariates

Dependent variables were the perceived trustworthiness of the respective news items, the willingness to share, interact, and forward the news item to a friend or family member, measured on a 7-point Likert scale, ranging from 1 = Very untrustworthy / very unlikely to 7 = Very trustworthy / very likely. All four behavioural responses were measured with seven stimulus items each.

Additionally, political knowledge was measured using four questions about the US political system: 1.) Which political office does Mike Pence currently hold? 2.) Whose responsibility is it to determine if a law is constitutional or not … is it the president, the Congress, or the Supreme Court? 3.) How much of a majority is required for the U.S. Senate and House of Representatives to override a presidential veto? 4.) Which party is more conservative (Democratic Party vs. Republican Party)? Correct answers were coded as 1, wrong answers as 0, and an overall score was computed.

Political orientation, news consumption, social media assessment

Participants were furthermore asked to assess their political orientation on a 7-point Likert scale from 1 = Very left to 7 = Very right, how frequently they consume news, either in print, online, or on TV, ranging from 'never' to 'several times a day'. Additionally, participants were asked about their preferred sources of news (television, online newspapers / magazines, print newspapers, social media, radio).

Furthermore, all participants were asked to assess the trustworthiness of all social media platforms used in the study, using a 5-point Likert scale (1 = very untrustworthy, 5 = very trustworthy).

Attention checks

Two attention checks were administered: after answering questions on the seven news items, every participant was asked from which social media platform the screenshots were taken. Correct answers were coded 1, wrong ones 0. At the end of the survey, participants were asked to briefly summarize what they think is the main purpose of this survey. All answers that provided some idea about the study's purposes were coded 1.

Answers that showed that the question was not read properly, e. g. by providing feedback about the study such as "good", "nice survey", etc. were coded as 0. Both variables were added to an overall attention score.

Recruitment and predetermined sample size

Participants were recruited on Amazon Mechanical Turk (MTurk) between August 31 and September 1, 2020 and received $1.80 for their participation.

Sample size was estimated using G*Power (Faul et al., 2007). Assuming a small effect size of f = 0,2, and a power = 0.99 (numerator df = 7, number of groups = 16, covariates = 2), the resulting minimum sample size to identify an interaction effect is N = 739.

Procedure

At the beginning of the survey, participants were asked about their consent to participate in the survey and informed about the length of the survey, about data anonymisation and intended use of the data.

After participants gave their informed consent, they were randomly assigned to one of the platform conditions. Afterwards, they were randomly presented with seven news items and asked to assess the trustworthiness of the item, their willingness to share, interact, and forward it to a friend or family member. After finishing these assessments an attention check was administered, asking participants about the platform where the news items were taken from. Participants were then asked to indicate the frequency with which they use the respective social media platform, the degree to which they consider misinformation is a problem on the platform and how likely they think it is they shared misinformation via the platform.

After these condition specific questions were answered, participants were asked to assess the trustworthiness of all seven social media platforms used in the survey and indicate whether they perceived social media to be more trustworthy than traditional media, such as TV and newspapers. Following this, questions on political orientation and political knowledge were administered before demographical characteristics were assessed. Lastly, participants were asked to briefly describe what they thought the purpose of the survey was. The survey ended with a debriefing, also offering contact information.

4. Results

The following section describes the descriptive results and inferential analyses that were conducted.

Demographics

Overall, 855 participants finished the survey, with age ranging from 18 to 72 (M = 33.45, SD = 10.79). Most participants were male (58.8%, n = 440). Participants were highly educated, with 65.5 percent (n = 560) having graduated from a university. Further 14.4 percent of participants (n = 123) had at least attended a university, but without receiving a degree.

Trustworthiness of social media platforms

Concerning the trustworthiness of social media platforms, LinkedIn descriptively appears to be the most trusted network, followed by Telegram, Twitter, Discord, Instagram, WhatsApp and lastly Facebook. Table 1 shows the mean trustworthiness of these platforms. No statistically significant differences between platforms were found

Table 1: Perceived trustworthiness of social media platforms

	N	M	SD
Facebook	854	2.61	1.29
WhatsApp	846	2.86	1.22
Instagram	851	2.92	1.21
Discord	852	2.93	.96
Twitter	850	3.03	1.30
Telegram	853	3.04	1.02
LinkedIn	854	3.57	1.01

Note. Trustworthiness was assessed on a 5-point Likert scale, from 1 = very untrustworthy to 5 = very trustworthy.

Comparing behavioural responses across platforms

Table 2 shows the number of participants in each group, as well as the means and standard deviations for all four assessed behavioural responses. Concerning trust, mean trust is highest for Twitter (M = 4.11, SD = 1.39), and lowest for WhatsApp (M = 3.93, SD = 1.30).

For sharing, people report the highest willingness to share news from LinkedIn (M = 3.47, SD = 1.50), and the lowest for Discord (M = 3.02, SD = 1.71). Concerning the interaction with the stimulus material, respondents were most likely to interact with news items from Facebook (M = 3.35, SD = 1.77), and least likely with material from Discord (M = 2.90, SD = 1.70). Forwarding news items was most likely when they were framed as coming from LinkedIn (M = 3.74, SD = 1.49) and list likely from Discord (M = 3.33, SD = 1.77).

Table 2: Behavioural responses to news items by platform

Platform		Trust	Sharing	Interacting	Forwarding
	N	M (SD)	M (SD)	M (SD)	M (SD)
Discord	101	3.99 (1.27)	3.02 (1.71)	2.90 (1.70)	3.33 (1.77)
Facebook	108	4.09 (1.25)	3.37 (1.75)	3.35 (1.77)	3.52 (1.79)
Instagram	107	3.94 (1.40)	3.27 (1.68)	3.14 (1.71)	3.39 (1.79)
LinkedIn	100	4.04 (1.21)	3.47 (1.50)	3.28 (1.58)	3.74 (1.49)
Telegram	110	4.05 (1.15)	3.29 (1.60)	3.11 (1.64)	3.52 (1.59)
Twitter	108	4.10 (1.18)	3.36 (1.65)	3.30 (1.66)	3.71 (1.54)
WhatsApp	112	3.93 (1.30)	3.10 (1.61)	3.08 (1.68)	3.34 (1.69)
Associated Press	109	4.01 (1.14)	3.21 (1.64)	3.15 (1.57)	3.47 (1.59)
Total	855	4.02 (1.24)	3.26 (1.64)	3.16 (1.66)	3.50 (1.65)

Note. All variables assessed on a 7-point Likert scale, with 1 = very untrustworthy / very unlikely to 7 = very trustworthy / very likely.

An analysis of covariances (ANCOVA) was conducted to identify the effects of platform on perceived trustworthiness of the news items and behavioural responses to them, using political knowledge as covariate.[39] The effect of platform on trustworthiness [F(7, 837) = .474, p = .854, partial n2 = .004], willingness to share [F(7, 837) = .947, p = .470, partial n2 = .008], willingness to interact with the news items [F(7, 837) = .789, p = .597, partial n2 = .007], and the willingness to forward the items [F(7, 837) = 1.066, p = .383, partial n2 = .009] all failed to reach statistical significance. In other words, the differences reported in Tables 1 and 2 are not significant.

[39] The second covariate, frequency of use, had to be dropped from the design described in the pre-registration due to violations of assumptions.

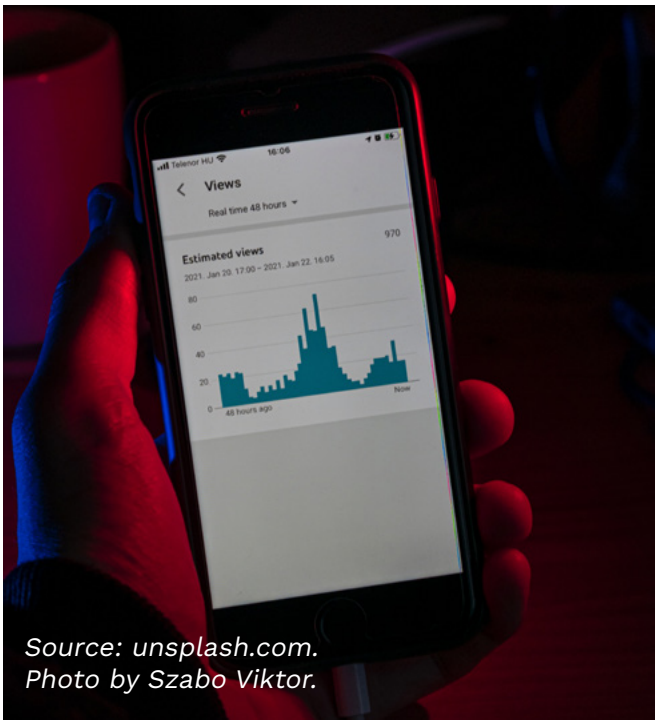
Exploratory analysis

Other factors than superficial characteristics of SNS seem to matter for trustworthiness and individual behaviour in relation to news. It may be that people who do not find their views reflected in traditional media perceive social media as an alternative territory, invest it with more credibility and care little about channels and their reputation, in particular if they perceive controversy to be the price for espousing alternative views marginalized in traditional media. To further understand which factors drive the perceived trustworthiness of news items and seem to act equally across SNS channels, I conducted an additional exploratory analysis.

This analysis is significant in view of public policies against misinformation which could be developed. Several authors have identified countermeasures that could be directly implemented on the SNS interface by technology companies to cope with the persistent problem of misinformation. Warning labels, intended to increase transparency, have been frequently employed by different SNS. In 2018, YouTube started to implement a label below its videos, identifying state-affiliations of the channel that uploaded the video (Samek, 2018), although inconsistently (Kofman, 2019). Facebook similarly highlighted a site's state-affiliation (Rosen et al., 2019) and extended its policy to include misleading information about COVID-19 (Rosen, 2020), which is also frequently updated. Twitter has a similar policy for identifying information distributed via tweets (Twitter, 2020).

Research has provided evidence for labels' effectiveness to mitigate the effects of misinformation. Arnold et al. (2021) found that source-related alerts, which inform users about the source of pseudonymous content posted on social media, e.g., by Russian government affiliates, reduce belief in the content's message and the users' likelihood to further spread the content. Although they found that the effects are influenced by partisanship, social media type, and the alert's specificity ("foreign government" vs. "Russian government"). The practice of highlighting media outlets' state-affiliation through warning labels could prevent opinion change (Nassetta & Gross, 2020).

The effects of prior exposure to misinformation on the platforms, whether participants have shared misinformation before, their political orientation (left vs. right leaning), frequency of use, news consumption, perceived credibility of traditional vs. social media, preferred news source (television, newspapers online and print vs. social media), political knowledge, as well as demographic variables (gender, education, age) were of great interested and therefore separately analysed. Participants' score in the attention check was also included in the model. Table 3 shows the results of the regression analysis.



Source: unsplash.com.
Photo by Szabo Viktor.

Table 3: Summary of regression for variables predicting perceived trustworthiness

Variable	Model 1			Model2			Model 3		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Misinformation	.18**	.03	.03	.15**	.025	.18	.14	.03	.17
Misinformation shared	.21**	.02	.02	.16**	.022	.25	.16	.02	.25
Credibility SM vs. TM	-.17**	.03	.03	-.12**	.03	-.11	-.12**	.03	-.11
Political orientation				.06	.02	.08	.07*	.03	.09
Political knowledge				-.10**	.02	-.11	-.10**	.03	-.11
Frequency of use				.08	.04	.07	.08	.04	.06
News source				.08	.08	.03	.08	.08	.03
Attention check				-.35**	.09	-.13	-.35**	.09	-.13
News consumption				.06*	.02	.07	.05*	.02	.06
Gender							.07	.07	.03
Education							.06*	.02	.08
Age							-.00	.00	-.01
R2	.27			.30			.31		
F for change in R2	99.08**			7.37**			2.49		

*Note. Dependent variable is the news items' perceived trustworthiness. N = 855. *p < .05 **p < .01*

The exploratory analyses revealed several substantial effects. Participants who perceive the SNS to have a misinformation problem rate the trustworthiness of news items presented on that platform significantly higher. Furthermore, if participants have shared misinformation before, they are more likely to judge the news as trustworthy. Those who find social media more credible than traditional media also assess the news more favourably. Political knowledge is furthermore negatively associated with credibility judgements.

News consumption, ranging from never to several times a day, shows a positive relationship with trustworthiness assessments. Participants' score in the attention check manipulation shows a negative association with the perceived news trustworthiness. Whether participants prefer social media over traditional media as sources of their news, does not impact their trustworthiness perceptions. Demographic variables do not offer any explanatory value, except for education.

5. Conclusion

This paper attempted to investigate the impact of SNS on a set of behavioural responses relevant to the spread of misinformation on social media.

Several different SNS, as well as the AP website, a wide range of different social media platforms – from messenger services like Telegram and WhatsApp, to gaming community site Discord and business network LinkedIn – were included in this investigation.

Contrary to the initial hypothesis, there does not seem to be a genuine platform effect – none of the variables of interest differ across the SNS. These results are somewhat counterintuitive. It could be expected that news coming from the AP website is perceived to be more trustworthy, and are therefore more likely to elicit behavioural responses, than news from a rather unknown network such as Discord. Beyond that, a business network such as LinkedIn, which has so far not been at the centre of a large-scale misinformation scandal, should be substantially more trustworthy to respondents. Yet, it does not seem to be more likely to make users spread its content than other networks.

Even platforms like Telegram and WhatsApp, which have a scandal-ridden track record, do not affect participants' trustworthiness assessments or their behavioural responses. Given that Telegram is among the preferred platforms of both QAnon conspiracists – as evidenced by its use among QAnon conspiracists in the storming of the United States Capitol Building during the January 6th insurrection (Rogers, 2020) - and terrorist group ISIS – with ISIS especially recommending the platform to its terrorists due to the service's reputation for offering high encryption standards (Weimann, 2016), these findings are somewhat surprising. Even the high salience of the ongoing Infodemic (World Health Organization, 2020), during which WhatsApp has become a preferred tool for misinformation agents in their attempt to spread false and misleading claims about the ongoing COVID-19 pandemic, does not elicit more cautious behaviour when confronted with suspicious news.

Therefore, in contrast to existing evidence of SNS-specific effects (Bode & Vraga, 2018; Stockmann et al., 2020) the presented findings do not support these notions. However, the exploratory analysis did reveal some interesting findings. The less participants see misinformation as a problem on a SNS, the more likely they are to assess the news in a favourable way. In contrast, if participants had shared misinformation before, they tend to perceive the news to be more credible. While the first association is intuitively clear, the latter seems surprising. One would expect that those who had likely shared misleading claims in the past excerpt more caution when asked to assess the trustworthiness of news. The opposite seems instead to be the case.

The results furthermore indicate that political orientation has positive association with higher trustworthiness judgements. This finding is unsurprising as a partisan bias is well documented in the literature (Bullock et al., 2015; Taber & Lodge, 2006; Van Bavel & Pereira, 2018), since all stimuli presented in this survey rather buy into right-wing conspiracy theories. Beyond that, a higher political knowledge is associated with lower trustworthiness ratings, also largely aligned with existing literature.

Especially assumptions made by dual-process models suggest that shallow information processing is dominant where little motivation or resources exist to thoroughly assess information (Kahnemann, 2011; Shiffrin & Schneider, 1977). Higher political knowledge makes credibility assessments easier, which protects users from falling for false or misleading information that appears credible given its presentation.

In the face of these findings, it becomes clear that further research is needed to address this apparent mismatch of public scandals and trust in as well as the willingness to engage with news presented on these platforms. Additionally, further research should investigate ways to make SNS misinformation problems more salient to users as an approach to mitigate misinformation. Warning labels, as used by Facebook, Twitter, and YouTube, could be adapted to inoculate users against misinformation, by showing how frequent an SNS has been involved in disseminating misinformation over a certain period, thereby increasing users' awareness of the severity of the problem. This could serve as a simple inoculation strategy that simultaneously strengthens public accountability of SNS as their inability or unwillingness to constrain the spread of falsehoods is on display.

Furthermore, research in this field should pay more attention to the various forces individuals are exposed to when using the internet or SNS. As described in the analytical model, individuals are not only affected by intra-person factors and their immediate physical environment, but also by situational factors that affect and are affected by their use of digital technology, stress, anxiety, and other emotional states. It is therefore crucial to further dissect these differing layers of interconnected factors to get a comprehensive picture of how misinformation is influencing individuals and what we can do to increase not only individual but also societal resilience.

6. Limitations

Assessing the willingness to interact with screenshots of course offers a different experience than engaging with the actual social media platform. Social media is designed to gain and keep to attention of its users in a way that is difficult (if at all) to replicate in a research environment. It is therefore reasonable to assume that directly assessing behaviour while people browse on social media would be a more natural approach. Furthermore, the screenshots presented in this paper showed the web applications of the respective platforms. There is, however, a substantial number of users interacting with social media through the platforms' mobile apps. Mobile apps might have a different effect on behavioural outcomes than their web-based counterparts. Future research should address this issue and identify potential differences in web and mobile applications.

Cultural aspects of the SNS are also not addressed in this study. It is possible that certain norms for how a SNS is used emerge over time, making a network more or less likely to host misinformation. It is imaginable, for example, that the business context in which LinkedIn is used normatively restrains the willingness of users to discuss seemingly conspiratorial claims, such as PizzaGate or President Obama's falsely claimed 'missing' birth certificate.

In contrast, Telegram's attraction to right-wing extremist groups, such as QAnon conspiracists, might create the impression that this type of content is welcome on the platform and community. Aside from the structural differences discussed in this study, it is therefore possible that behavioral norms emerge on different SNS over time, depending on the kind of users the platforms manage to attract. That might be especially the case for some SNS that are initially selected by users for very particular reasons – Telegram, for example, was initially chosen by many for its high encryption standards.

Lastly, the significant number of participants which have failed one attention check could also bias the results. As roughly one third (31.3%) of respondents has answered only one out of two checks correctly, it is not clear whether an increased cognitive effort could lead to different results. However, the highly significant negative association of the attention check score with the perceived trustworthiness of the news items could be interpreted as support for biased results.

References

Abalakina-Paap, M., Stephan, W. G., Craig, T., & Gregory, W. L. (1999). Beliefs in conspiracies. Political Psychology, 20(3), 637–647. https://doi.org/10.1111/0162-895X.00160

Arnold, J. R., Reckendorf, A., & Wintersieck, A. L. (2021). Source alerts can reduce the harms of foreign disinformation. Harvard Kennedy School Misinformation Review, 1(May), 1–33. https://doi.org/10.37016/mr-2020-68

Bode, L., & Vraga, E. K. (2018). Studying Politics Across Media. Political Communication, 35(1), 1–7. https://doi.org/10.1080/10584609.2017.1334730

Boyd, danah m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). Types, Sources, and Claims of COVID-19 Misinformation. In Oxford University Press (Issue April).

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2015). Partisan Bias in Factual Beliefs about Politics. Quarterly Journal of Political Science, 10(4), 519–578. https://doi.org/10.1561/100.00014074

Burnes, B. (2007). Kurt Lewin and the harwood studies: The foundations of OD. Journal of Applied Behavioral Science, 43(2), 213–231. https://doi.org/10.1177/0021886306297004

Burnes, B., & Cooke, B. (2013). Kurt Lewin's field theory: A review and re-evaluation. International Journal of Management Reviews, 15(4), 408–425. https://doi.org/10.1111/j.1468-2370.2012.00348.x

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 39(5), 752–766. https://doi.org/10.1037//0022-3514.39.5.752

Chuai, Y., & Zhao, J. (2020). Anger makes fake news viral online. http://arxiv.org/abs/2004.10399

Diamond, G. A. (1992). Field Theory and Rational Choice: A Lewinian Approach to Modeling Motivation. Journal of Social Issues, 48(2), 79–94. https://doi.org/10.1111/j.1540-4560.1992.tb00885.x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. New Media and Society, 9(2), 319–342. https://doi.org/10.1177/1461444807075015

Flanagin, A. J., Winter, S., & Metzger, M. J. (2020). Making sense of credibility in complex information environments: the role of message sidedness, information source, and thinking styles in credibility evaluation online. Information Communication and Society, 23(7), 1038–1056. https://doi.org/10.1080/1369118X.2018.1547411

Garimella, K., & Eckles, D. (2020). Images and Misinformation in Political Groups: Evidence from WhatsApp in India. Harvard Kennedy School Misinformation Review, 1(5), 1–12. https://doi.org/10.37016/mr-2020-030

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. Nature Human Behaviour, 4(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Hampton, K. N., Lee, C., & Her, E. J. (2011). How new media affords network diversity: Direct and mediated access to social capital through participation in local social settings. New Media & Society, 13(7), 1031–1049. https://doi.org/10.1177/1461444810390342

Imhoff, R., & Bruder, M. (2014). Speaking (Un-)truth to power: Conspiracy mentality as a generalised political attitude. European Journal of Personality, 28(1), 25–43. https://doi.org/10.1002/per.1930

Isenberg, D. J. (1986). Group Polarization. A Critical Review and Meta-Analysis. Journal of Personality and Social Psychology, 50(6), 1141–1151. https://doi.org/10.1037/0022-3514.50.6.1141

Johnson, H. M., & Seifert, C. M. (1994). Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Kahnemann, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.

Kang, H., Bae, K., Zhang, S., & Sundar, S. S. (2011). Source cues in online news: Is the proximate source more powerful than distal sources? Journalism and Mass Communication Quarterly, 88(4), 719–736. https://doi.org/10.1177/107769901108800403

Kofman, A. (2019, November). YouTube promised to label state-sponsored videos but doesn't always do so. ProPublica. https://www.propublica.org/article/youtube-promised-to-label-state-sponsored-videos-but-doesnt-always-do-so

Lazer, D., Baum, M., Benkler, J., Berinsky, A., Greenhill, K., Metzger, M., Nyhan, B., Pennycook, G., Rothschild, D., Sunstein, C., Thorson, E., Watts, D., & Zittrain, J. (2018). The science of fake news. Science, 359(6380), 1094–1096. https://doi.org/https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Cook, J., Oberauer, K., Brophy, S., Lloyd, E. A., & Marriott, M. (2015). Recurrent fury: Conspiratorial discourse in the blogosphere triggered by research on the role of conspiracist ideation in climate denial. Journal of Social and Political Psychology, 3(1), 142–178. https://doi.org/10.5964/jspp.v3i1.443

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. Psychological Science in the Public Interest, Supplement, 13(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA Faked the Moon Landing-Therefore, (Climate) Science Is a Hoax: An Anatomy of the Motivated Rejection of Science. Psychological Science, 24(5), 622–633. https://doi.org/10.1177/0956797612457686

Lewin, K. (1939). Field Theory and Experiment in Social Psychology: Concepts and Methods. American Journal of Sociology, 44(6), 868–896. https://www.jstor.org/stable/2769418

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology, 58(13), 2078–2091. https://doi.org/10.1002/asi.20672

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. Journal of Communication, 60(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Meyer, M., Alfano, M., & de Bruin, B. (2021). Epistemic vice predicts acceptance of Covid-19 misinformation. Episteme, 1–22. https://doi.org/10.1017/epi.2021.18

Mutz, D. C., & Martin, P. S. (2001). Facilitating Communication across Lines of Political Difference: The Role of Mass Media. American Political Science Review, 95(1), 97–114.

Nassetta, J., & Gross, K. (2020). State media warning labels can counteract the effects of foreign disinformation. Harvard Kennedy School Misinformation Review, 1(October), 1–11. https://doi.org/10.37016/mr-2020-45

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188(June), 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pezzo, M. V., & Beckstead, J. W. (2006). A multilevel analysis of rumor transmission: Effects of anxiety and belief in two field experiments. Basic and Applied Social Psychology, 28(1), 91–100. https://doi.org/10.1207/s15324834basp2801_8

Rains, S. A., & Brunner, S. R. (2015). What can we learn about social network sites by studying Facebook? A call and recommendations for research on social network sites. New Media and Society, 17(1), 114–131. https://doi.org/10.1177/1461444814546481

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. Annual Review of Information Science and Technology, 41, 307–364. https://doi.org/10.1002/aris.2007.1440410114

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. European Journal of Communication, 35(3), 213–229. https://doi.org/10.1177/0267323120922066

Rosen, G. (2020). An update on our work to keep people onformed and limit misinformation about COVID-19. Facebook. https://about.fb.com/news/2020/04/covid-19-misinfo-update/

Rosen, G., Harbath, K., Gleicher, N., & Leathern, R. (2019). Helping to protect the 2020 US Elections. Facebook. https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/

Rossini, P., Stromer-Galley, J., Baptista, E. A., & Veiga de Oliveira, V. (2020). Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections. New Media and Society. https://doi.org/10.1177/1461444820928059

Samek, G. (2018). Greater transparency for users around news broadcasters. YouTube. https://blog.youtube/news-and-events/greater-transparency-for-users-around

Scott, M. (2020, March 30). Facebook's private groups are abuzz with coronavirus fake news. POLITICO. https://www.politico.eu/article/facebook-misinformation-fake-news-coronavirus-covid19/

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review, 84(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J., & Loker, K. (2019). Who Shared It?: Deciding What News to Trust on Social Media. Digital Journalism, 7(6), 783–801. https://doi.org/10.1080/21670811.2019.1623702

Stockmann, D., Luo, T., & Shen, M. (2020). Designing authoritarian deliberation: how social media platforms influence political talk in China. Democratization, 27(2), 243–264. https://doi.org/10.1080/13510347.2019.1679771

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), Digital Media, Youth, and Credibility (pp. 73–100). The MIT Press. https://doi.org/10.1162/dmal.9780262562324.073

Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. Journal of Communication, 51(1), 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. Applied Cognitive Psychology, 24(6), 749–761. https://doi.org/10.1002/acp.1583

Swami, V., & Furnham, A. (2014). Political paranoia and conspiracy theories. In Power, politics, and paranoia : why people are suspicious of their leaders. https://doi.org/10.1017/CBO9781107415324.004

Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation: Challenges and recommendations. Annual Review of Public Health, 41, 433–451. https://doi.org/10.1146/annurev-publhealth-040119-094127

Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. American Journal of Political Science, 50(3), 755–769.

Taber, C. S., & Young, E. (2013). Political Information Processing. In L. Huddy, D. O. Sears, & J. S. Levy (Eds.), The Oxford Handbook of Political Psychology (pp. 525–558). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199760107.001.0001

Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. Communications of the ACM, 42(5), 39–44. https://dl.acm.org/doi/fullHtml/10.1145/301353.301402?casa_token=h24p3CCfU4wAAAAA:isWZh5r4rBAW9CxjyDq69iRxnW-isH9J9ZnzoYzL8mVc-zdrLqt-9zLoWV8efhhZv9cOh-jhgE0I7

Twitter. (2020). Expanding our policies to further protect the civic conversation. https://blog.twitter.com/en_us/topics/company/2020/civic-integrity-policy-update.html

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: a challenge for democracy? Annals of the International Communication Association, 41(1), 3–27. https://doi.org/10.1080/23808985.2017.1288551

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. Trends in Cognitive Sciences, 22(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Vraga, E. K., & Bode, L. (2017). I do not believe you: how providing a source corrects health misperceptions across social media platforms. Information, Communication & Society, 21(10), 1337–1353. https://doi.org/10.1080/1369118X.2017.1313883

Weimann, G. (2016). Terrorist Migration to the Dark Web. Perspectives on Terrorism, 10(3), 40–44. https://www.jstor.org/stable/26297596

World Health Organization. (2020). Novel Coronavirus (2019-nCoV): Situation Report - 13. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf

Appendix

Appendix 1: Analysis of covariates

Political knowledge and platform usage were intended to serve as covariates in the model. To fulfil the criterion of independence from treatment, both covariates were analysed to using an ANOVA approach. Table 4 shows the descriptive statistics for both covariates. The ANOVA revealed a significant platform difference for frequency of use $F(7, 839) = 19.37$, $p = .92$. Frequency of use could therefore not be included in the ANCOVA. No differences are found for political knowledge $F(7,845) = 1.74$, $p = .09$ and is hence included in the ANCOVA.

Table 4: Descriptive statistics for covariates

Platform		Political Knowledge	Frequency of use
	N	M (SD)	M (SD)
Discord	101	2.49 (1.22)	2.03 (1.01)
Facebook	108	2.28 (1.30)	3.04 (0.90)
Instagram	107	2.23 (1.34)	2.88 (0.09)
LinkedIn	100	2.22 (1.32)	2.45 (0.87)
Telegram	110	2.50 (1.19)	1.93 (0.98)
Twitter	107	2.32 (1.38)	2.60 (0.89)
WhatsApp	112	2.40 (1.31)	2.70 (1.16)
AP	109	1.99 (1.30)	2.10 (0.92)
Total	854	2.30 (1.30)	2.47 (1.03)

Note. All variables assessed on a 7-point Likert scale, with 1 = very untrustworthy / very unlikely to 7 = very trustworthy / very likely.

Appendix 2: Item-level descriptive statistics

Table 5 shows the item-level statistics for all dependent variables, averaged across platforms.

Table 5: Item-level descriptive statistics

Item				Skewness	
Trust	**N**	**M**	**SD**	**Statistic**	**SE**
BLM	855	4.27	1.84	-.363	.084
RBG	855	4.28	1.64	-.381	.084
BW	855	3.74	1.85	.038	.084
Gates	855	3.93	1.82	-.154	.084
HO	855	3.93	1.86	-.149	.084
USPS	855	4.00	1.78	-.154	.084
Sharing	**N**	**M**	**SD**	**Statistic**	**SE**
BLM	855	3.41	2.01	.188	.084
RBG	855	3.35	1.93	.231	.084
BW	855	3.17	2.05	.369	.084
Gates	855	3.25	1.99	.310	.084
HO	855	3.23	2.04	.309	.084
USPS	855	3.24	1.99	.303	.084
Biden	855	3.17	2.01	.371	.084

Interacting	N	M	SD	Statistic	SE
BLM	855	3.15	2.02	.385	.084
RBG	855	3.28	1.91	.277	.084
BW	855	3.11	2.00	.406	.084
Gates	855	3.22	2.04	.360	.084
HO	855	3.09	2.00	.432	.084
USPS	855	3.21	2.01	.365	.084
Biden	855	3.09	2.02	.451	.084
Forwarding					
BLM	855	3.70	2.07	.002	.084
RBG	855	3.49	1.99	.131	.084
BW	855	3.45	2.12	.178	.084
Gates	855	3.49	2.09	.158	.084
HO	855	3.52	2.09	.121	.084
USPS	855	3.50	2.05	.132	.084
Biden	855	3.37	2.09	.252	.084

Call for Papers from the Romanian Journal of Political Science (PolSci)

Volume 21, N° 2, Winter 2021

PolSci (Romanian Journal of Political Science) is a bi-annual journal edited by the Romanian Academic Society (RAS). PolSci is indexed by ISI Thomson under the Social Sciences Citation Index and by other prominent institutions such as IPSA, GESIS, EBSCO, CEEOL and ProQuest. Previous PolSci contributors include Francis Fukuyama, Larry Diamond or Philippe Schmitter, as well as a significant number of young researchers. The Journal welcomes academic papers, reviews of recent publications and announcements of forthcoming volumes as well as contributions from various fields of research in the social sciences.

Submissions are now being accepted for the Winter 2021 issue, and subsequent volumes.

Please submit your work saved as a Microsoft Word document at **polsci@sar.org.ro**. Submissions are received on a rolling basis. Expected turnover is of 60 days.

Kindly note that we will not review manuscripts that have already been published, are scheduled for publication elsewhere, or have been simultaneously submitted to another journal.

The Romanian Academic Society requires the assignment of **copyright.** If you are interested in submitting an article, please visit our website for details on the **Authors' Guidelines**.

Managing Editor

Tel: 0040 212 111 424
E-mail: polsci@sar.org.ro

ISSN 1582-456X

PolSci
Romanian Journal of Political Science

Volume 21, Nº 1, Summer 2021