Experimental evidence of sharing and reporting misinformation in Switzerland, France, and Germany

Final Report Submitted to the Federal Office of Communication

Achim Edelmann and Christian Müller

11 February 2023

The digital transformation has led to major changes in the information ecosystem. Modern communication platforms such as WhatsApp, Facebook and Telegram have become important media for information dissemination, blurring boundaries between private and public information channels. Unfortunately, these platforms have also been misused to share inaccurate or inflammatory content. However, studying how misinformation spreads through and across such platforms is notoriously difficult due to scant data. This severely limits an understanding of how misinformation spreads in contemporary societies. Building on a web-service technology we have developed to track sharing of information, we implemented a series of controlled field experiments to research how information veracity and political thinking influences participants' sharing on Personal Messaging-based Platforms. To better understand how participants evaluate information in those decisions, we used non-intrusive behavioral measures that capture whether they report information as fake. We fielded experiments in the French- and German-speaking parts of Switzerland and compared results with experiments run in Germany and France. Across all countries, we find no difference in sharing of true and untrue information; at the same time, we find evidence that participants are able to identify and willing to report false information without being asked to do so. This suggests that participants do not share false information at the same rate as true information because they were not able to discern the difference between the two.

Introduction

Distributed and decentralized communication networks have become an essential part of the modern communication systems. This especially concerns Personal Messaging-based Platforms (PMP) such as Facebook Messenger, WhatsApp, or Telegram. Through their capacity to link information from and feed it back into more classic information streams, they increasingly influence the flow of communication. Unfortunately, however, little is known on how information spreads across PMPs. This is especially problematic since we know that PMPs are key channels for the spread of misinformation.

There are two major difficulties in studying how misinformation is shared across such networks. First, while methods to study how fake news is shared on social networking sites such as Facebook or Twitter have been developed (e.g., Guess et al. 2020, Allcott & Gentzkow 2017), no comparable research exists for PMPs. The main reason for this is that the end-to-end encryption built into these services prevents examining network features of the spread of misinformation where individuals act as nodes and misinformation is passed forward one contact at a time.

Second, we lack causal evidence on factors shaping people's sharing behavior and related reasoning. Most existing research that asks how misinformation (compared to true information) is shared describes traces of sharing in publicly available data (e.g. Vosoughi et al. 2018, Friggeri et al. 2014). This cannot answer, for example, whether and how story veracity causally influences sharing behavior. The few works that have attempted to address such causal questions have remained limited to study effects on people's *intent to share* while asking for *opinions* about a given story (e.g. Buchanan 2020). But this approach falls short on external validity. Without measuring participants' *sharing behavior* we cannot know for sure whether they would actually do so. What is more, asking them to evaluate a story has limited bearing on how they would perceive it if left to their own devices, let alone whether they would bear the costs of actively reporting false information.

In this project, we address these shortcomings. We ask whether and how the veracity of a story shapes participants' sharing behavior as well as whether and how they are actively denouncing false information. Building on recent works, we also ask whether thinking about politics mediates related effects (e.g. Osmundsen et al. 2021, Pennycook & Rand 2019).

To answer these questions, we conducted four controlled online experiments manipulating information veracity and participants' political thinking. We fielded

these experiments in four regions—the German and French speaking part of Switzerland, Germany, and France. In addition to countering the strong focus on the United States in the literature, comparing results of these experiments provide leverage on differences across language and cultural context. This is important as misinformation is often tied to (especially political) issues prevalent in a particular cultural context such as a country. Switzerland, however, is unique in hosting several languages within one such context. Even if written in these different languages, we can assume that misinformation is largely understood the same way within Switzerland. This offers the possibility to study differences across languages independent of contextual differences. Likewise, comparing results from Switzerland with those from France and Germany, we can study differences across contexts while keeping language constant.

In the following, we begin by describing the experimental setup, including the development of the stimuli and issues of technical implementation. We then summarize the fielding of the experiments, describing participant recruitment and the analytical sample. Turning to our results, we focus on the effects of *information veracity* and *political thinking* on two outcomes: *sharing behavior* and *active reporting* of false information. To conclude, we point out possible limitations and implications of this study.

Experimental setup

To answer these questions, we set up a series of controlled online experiments that directed participants to a web site, displaying a variety of stories. To make sure that respondents would evaluate the information rather than the credibility of the source (Flanagin & Metzger 2007, Fogg et al. 2001, Hong 2006, Johnson & Kaye 2004), we decided to design the layouts of the individual web pages in ways resembling those of newspapers common in the respective four regions. This included *24 Heures, Neue Zürcher Zeitung, Le Monde, and Frankfurter Allgemeine*.

On these web pages, we implemented a series of sharing options in the form of commonly used sharing buttons for PMPs as well as a Copy Link button. To track the sharing of the web page across users, we built on and extended a web-service technology that we had developed as part of a Spark Grant of the Swiss National Science Foundation (#190250). This technology allows us to keep track of the path a web page takes as it diffuses within the contacts networks underlying PMPs, without having to resort to proprietary data. To make this technology suitable for this project,

several adjustments were necessary: First, we extended its capacity from tracking sharing across WhatsApp to also include Facebook Messenger and Telegram. Second, we went to great extent to develop and refine a technical solution that ensures that participants who visit the web page will be shown a debrief specific to each information. This debrief is triggered by a timeout tailored to the displayed information or by any sharing activity. This ensures, to the best possible, that participants will be debriefed no matter their actions taken during the experiment.

In addition to tracking sharing, we also provided commonly used buttons to leave a comment, which trigger a pop-up with the option to tick, among other things, that the information displayed is “inaccurate” as well as to leave a written comment. Figure 1 shows two examples of the web page used in the German- and French-speaking parts of Switzerland with the different sharing options alongside the pop-up. For our analysis, we take two kinds of actions taken on those web pages as measures of our dependent variables—whether they clicked on any of the available sharing options and whether they reported the information as inaccurate.

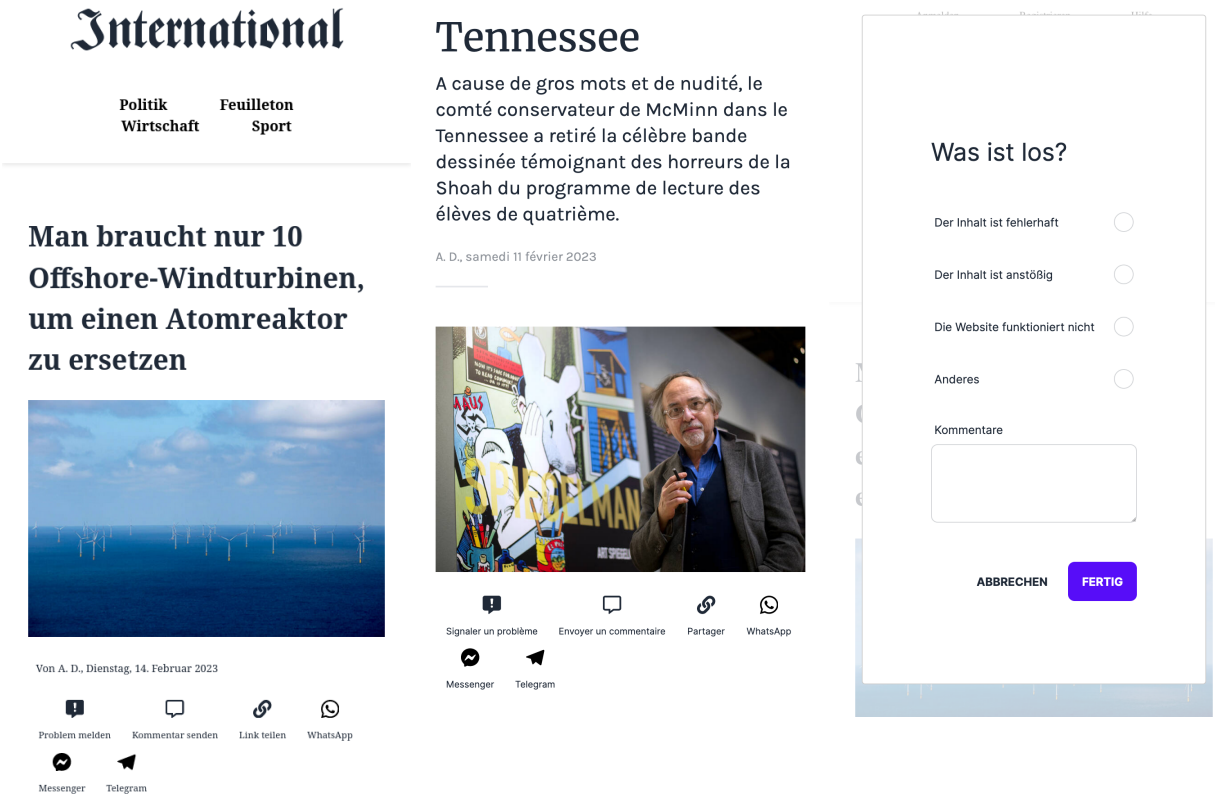

Figure 1: Screenshots of the web site showing two stories fielded in Switzerland in German (left) and French (center) with sharing and report buttons and the pop-up (right) to comment.

Political priming

To manipulate the first experimental variable—whether participants *think politically*—we randomly displayed half of the participants a brief pop-up survey upon entering the web site asking "With which party do you identify yourself?" (options were main political parties by country). This was followed by the question "How strongly do you identify yourself with this party?" (4 categories from "very strongly" to "very weakly").

Information veracity

We manipulated *story veracity* by selecting a set of true and false stories in German and French which we then adapted to the individual countries. To ensure external validity, we drew these stories from known fact-checking sites in Switzerland, France, and Germany (e.g. liberation.fr/checknews, correctiv.org, 20minutes.fr/societe/desintox, mimikama.org, politifact.com, faktencheck.afp.com, dpa-factchecking.com). We specifically searched for stories with a focus on events or salient topics in at least one of our target countries but could also easily be adapted to the other countries. For example, we chose a story comparing CO2 emissions of volcano eruptions against the yearly emission of cars over a story of whether Grenoble, the 2022 Green Capital of Europe, will produce all of its energy from renewable sources. While the story about green energy might be salient in France, it is very difficult to adapt it to Switzerland and Germany. Meanwhile CO2 emissions of each country—the comparison point to the (claimed) volcano emissions—is very easy to establish. The latter story only requires minor adjustments to make it appealing to participants in each country. Therefore, it is much more appropriate for this research project.

A second, important concern was that, despite our debriefing, none of the stories would pose risks to participants. This decision likely has bearings on our estimated sharing rates. Research has argued that new, unconventional, and emotionally arousing stories are more likely to be shared. Insofar as those characteristics are underrepresented in our stories, the sharing rates we report are to be considered as conservative estimates.

The process of carefully checking through all recent fact-checked stories and adjusting it to the appropriate context yielded 31 potential stories to include in the experiment. To ensure that the final stories were understood in the same way, we ran a pretest on Amazon MTurk in which we asked 88 participants to evaluate the

veracity and political salience of each story. Interestingly, the results in Figure 2 show that, overall, respondents were much more likely to rate stories as false, even if they are fact-checked as true. As the task specifically required MTurk workers to identify the truth of a story, the overall tendency to rate the stories as false is likely an artifact of the task because workers were already suspicious of every story they read. Nevertheless, we use these results to select five stories, which were best identified as true/false by respondents, i.e. false stories which were consistently identified as false and vice versa, because the experimental setup relies on participants being able to reliably detect the veracity of a story.

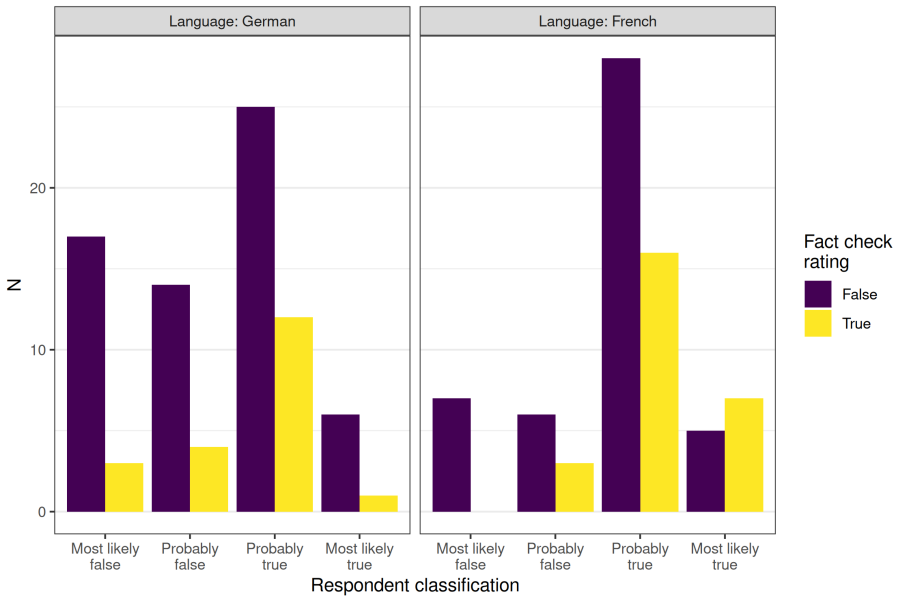

Figure 2: Classification of veracity for 31 stories in pretest.

Fielding of the experiments

Using the above setup, we fielded experiments in the German- and French-speaking regions of Switzerland as well as in Germany and France.

Participant recruitment

To recruit participants in these populations, we used Facebook Advertising. This offered several advantages. First, Facebook Advertising can target users based on

individual characteristics thus allowing us to recruit participants stratified across basic demographic characteristics, including age, gender, and education without collecting those from participants. Targeted advertising also ameliorates potential biases due to algorithmic advertising optimization and enables us to analytically control for participants' basic demographic characteristics. Second, it allows for optimizing adverts and spending based specific to target audiences. We used more than 12 different adverts tailored to different audiences and varied spending by category to ensure minimum subsamples in all combinations. Third, Facebook Advertising allows us to specify retargeting strategies as well as to exclude participants that have already clicked on one of our adverts. This helps to prevent duplicated participation and greatly eases removal of repeated participations and bots when defining the analytic sample (see below).

Despite these advantages, there are several disadvantages to our recruitment strategy. First, Facebook's user group is known to be biased. While targeted advertising reduces biases due to algorithmic advertising, our results only have external validity to the Facebook user base which may well differ across countries. Second, we needed to rely on Facebook's measurement of the targeting characteristics. These are likely less accurate than self-report. Third, costs for participant recruitment depends on the competitive edge of our adverts in a highly competitive attention market.

A major complication for our study was that costs of advertising were higher than expected. In all three countries, recruiting participants from certain categories was extremely costly (e.g. highly educated/young men). Presumably due to a higher standard of living but also due to smaller audience sizes, costs per participant reached up to 30 CHF in these categories. While this was surmountable in the French and German population, we needed to take additional measures in the Swiss-German and Swiss-French population. To obtain a reasonably sized sample, we therefore decided to collapse targeted audiences by gender. Nevertheless, average costs for participants in the Swiss-German and the Swiss-French population remained far above our estimates. Moreover, we also had to accept that in the French speaking part of Switzerland, we were not able to obtain a sample comparable in size to that in the German speaking part.

We worked with a reliable and efficient advertising company, which we had worked with before, to keep recruitment costs as manageable as possible. For example, with their help we could identify moments when to construct new adverts to

counter overall ad fatigue or to extend the recruitment window to avoid high stake competitions for users online which we did in all countries. However, their work also factors into the total costs. Combining all these efforts, we were able to keep the average cost per sign-up in our experiments under 10.- CHF. This is more than we had expected but still much less than what a big Swiss survey company quoted to use their sample.

Sample composition

We gathered a total of 3844 initial sign-ups through the advertisements over the four experiments. Not all of those sign-ups correspond to actual, unique participants. Although we used a Facebook Advertising mechanism to stop multiple participation, it is not perfect as it relies on cookies which many users restrict. Moreover, some participants visit the article multiple times or bots visit the article creating spurious "participants." We therefore employed several measures to ensure that the causal effects we estimate are not biased by those issues. These measures led to a smaller analytical sample and less statistical power, but ultimately better, more trustworthy results.

The steps to create a clean analytical sample comprised: (1) manually removing participants who repeatedly signed up through the ads, (2) removing repeated views of the article by the same participant, (3) filtering out views by bots, (4) geolocating the origin of views and only keeping the ones from the target countries, and (5) ensuring that only views from participants recruited through the advertising campaign remain in the analysis. Thanks to these precautions, we could identify 1030 visits to our web page that were either suspicious or unattributable. After their removal, we ended up with an analytical sample of 803 and 329 participants in the German- and French-speaking parts of Switzerland and 823 and 859 participants in Germany and France, respectively. Figure 3 shows the composition of this sample by age and education. Although the distributions are unequal and unrepresentative, this is easily corrected for in the analyses below.

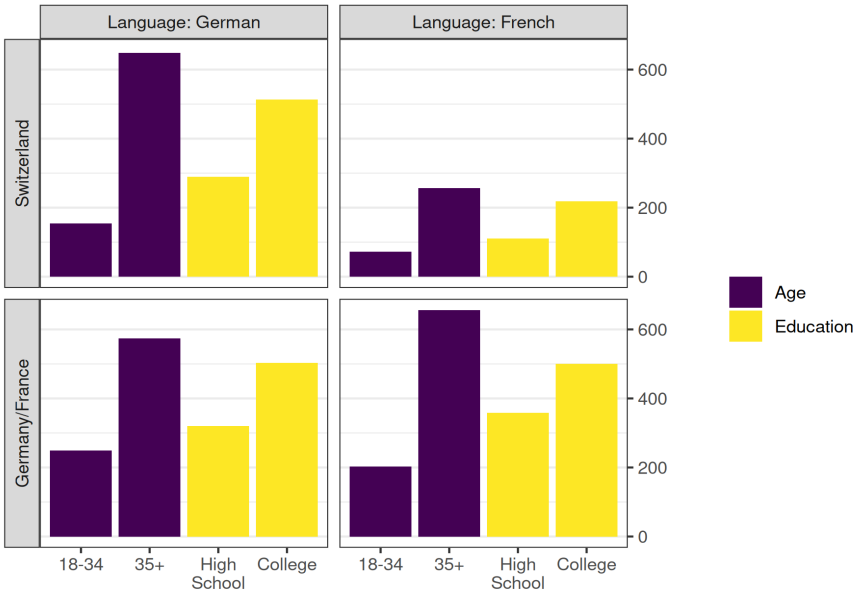

Figure 3: Composition of analytical sample.

Results

We fit two models to predict whether *participants share* and whether they *report the information as inaccurate* by the veracity of the story and whether they received a political primer. Turning to the results for participants' sharing, estimates for the base, unadjusted model (red points/lines) in Figure 4 suggest that the veracity of a story has no influence on how often it is shared. Estimated population sharing proportions for true stories without political priming are around 6% and 4% in the German- and French-speaking parts of Switzerland, respectively. None of the conditions show statistically significant differences from this base sharing rate, i.e. neither the veracity of the story nor the presence or absence of a politics prime plays a role in how often participants attempted to share the article. Adjusting for the covariates—age (18-34/35+) and education (less than high school/high school or more)—to increase the efficiency of the estimator (blue points/lines) does not change this result. The same applies to estimates for Germany and France in the lower part of Figure 4. While the point estimate for base sharing rates is slightly higher than in Switzerland, neither veracity nor political priming has a significant effect on sharing rates. Therefore, we conclude that false stories are not shared more or less often than true stories.

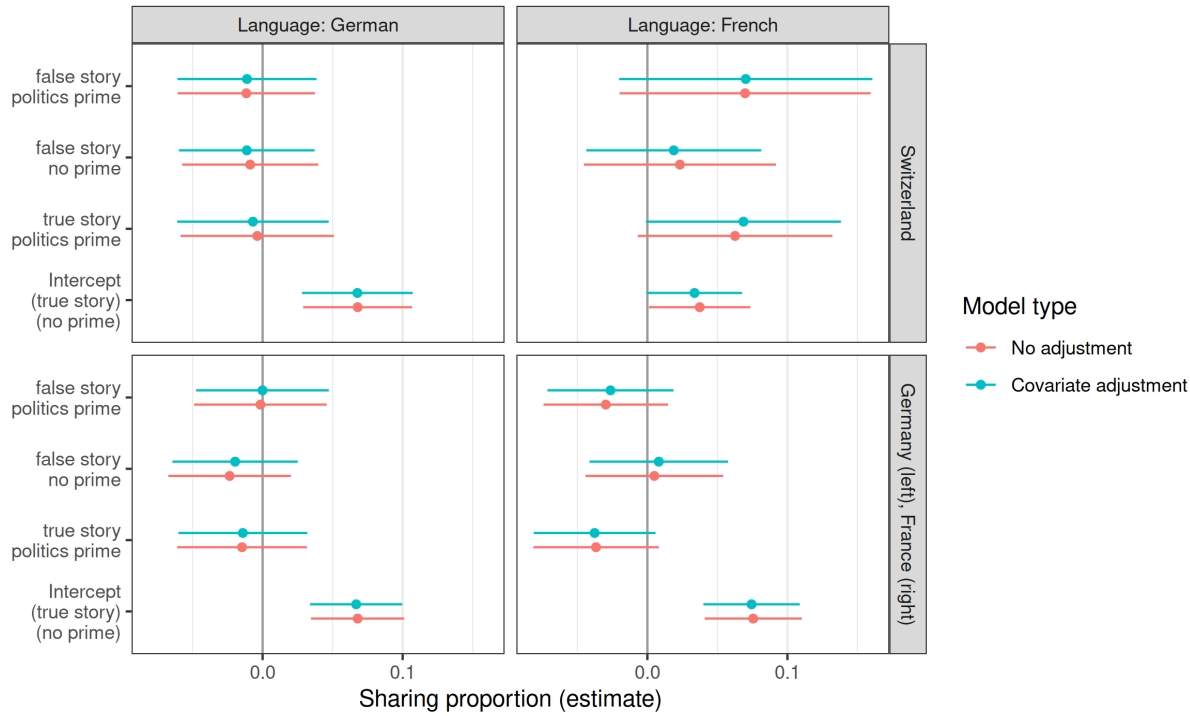

Figure 4: Estimated sharing proportions by treatment group and experimental sample.

We next ask whether participants are able to identify a story as false and are, crucially, also willing to take the time and effort to report that particular story is inaccurate, often accompanied by a comment addressing respective shortcomings. Results in Figure 5 show a completely different picture. All experimental groups, with the exception of German speakers in Switzerland, show clear, statistically significant differences between true and false stories. Specifically, the estimated base reporting proportions (i.e. the intercept for true stories without a politics prime) as well as the coefficients for the difference between true stories with a politics prime are not statistically significant. This suggests that true stories, independent of the priming condition, are, on average, not reported as false. At the same time, the coefficients for false stories are significant and positive (for politically primed and not primed) at around 3% to 7%, which shows that false stories are, on average, reported as inaccurate. The strengths of our design enables us to measure not only whether participants correctly identified a story as false but also whether they made an additional effort to report the story as false. Our results thus suggest that the population is actually able to discern true and false stories and some people are even willing to make unsolicited reports. Taken together, these two results suggest that participants do not share false information at the same rate as true information because they were not able to discern the difference.

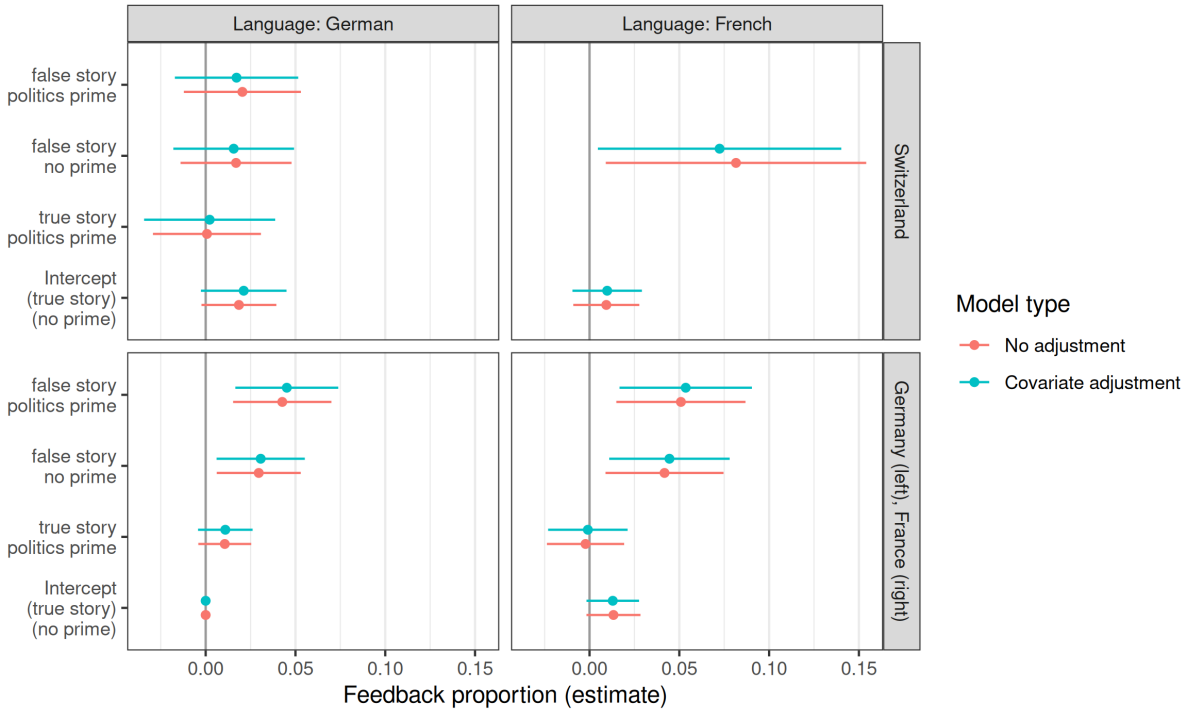

Figure 5: Estimated proportions of respondents indicating the article contains false information by treatment group and experimental sample; for France, pooled averages across no/politics prime shown due to too sample size.

Conclusion

The results in this field experiment paint a rather positive but also interesting picture of people's sharing behavior, digital literacy, and civil responsibility that raises new questions. On the one hand, false information online is not shared more or less often than true information; on the other hand, people, on average, are not only able to identify false information but also voluntarily bear the costs of reporting it without being asked to do so. These results are consistent across Switzerland, Germany, and France. Taken together, they suggest that it is not the inability to differentiate between true and false information or a general lack of motivation to act differently upon such perceptions that leads people to share both equally. This raises the question that, if it is not a lack of digital literacy nor willingness to act upon a noticed difference, why do people still share misinformation to the same extent as true information? And, relatedly, is increasing people's abilities to discern and willingness to report misinformation sufficient to curb its sharing?

Of course, as with any study results, also these results come with a number of limitations. First, while we took great care in assembling and pretesting stories used in the experiments, they are not a “representative sample” of false or true stories (let alone that it is conceptually hard to conceive of something like that) and our results are contingent on them. Second, due to issues of sample size our analyses might suffer from low statistical power (especially in the French-speaking part of Switzerland). This means that with larger samples we might have detected differences where we did not. Third, our sample selection hinges on our recruitment strategy. The Facebook user group is not representative of the general population in our target regions and Facebook’s advertising algorithms are (by definition) not random. This means that also our samples are not representative of any clearly definable population to which our results could be generalized.

Acknowledgements

We are extremely grateful for the financial support provided by the *Federal Office of Communication* and the *Berne University Research Foundation*, which allowed us to realize this ambitious project. All financial resources were used in accordance with the respective grant agreement. We also thank the Federal Office of Communication for the opportunity to present an early version of our research at the workshop “Digitale Desinformation und Hassrede” in April 2022 and its participants, including researchers, civil society representatives, and policy makers working on the issue of misinformation in Switzerland for their helpful contributions.

References

Allcott, H. & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211-36.

Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. Plos one, 15(10), e0239666.

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 & society, 9(2), 319-42.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., ... & Treinen, M. (2001, March). What makes web sites credible? A report on a large quantitative study. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 61-8).

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 101-110).

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

Hong, T. (2006). The influence of structural and message features on Web site credibility. Journal of the American Society for Information Science and Technology, 57(1), 114-27.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. American Political Science Review, 115(3), 999-1015.

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, 39-50.

Johnson, T. J., & Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions of weblogs among blog users. Journalism & Mass Communicatio Quarterly, 81(3), 622-42.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. science, 359(6380), 1146-1151.