

IFN Working Paper No. 1566, 2026

The Effect of Anonymity in Social Media on Fake News Sharing

Joakim Jansson and Emma von Essen

The Effect of Anonymity in Social Media on Fake News Sharing*

Joakim Jansson[†], Emma von Essen[‡]

August 31, 2026

Abstract

We study how anonymity affects fake-news sharing on social media using a natural experiment on Flashback, a large Swedish online discussion forum. In 2014, journalists obtained information that allowed them to identify users registered before March 2007, sharply increasing these users' risk of public exposure while leaving later registrants unaffected. Using a difference-in-differences design on 1.98 million posts, we find that exposed users reduce references to fake-news sites by about one-quarter of the mean. The effect is concentrated in immigration and domestic-politics discussions and does not extend to legitimate news sources, suggesting that accountability reduces misinformation without reducing news sharing more broadly.

JEL: D00, D80, D83, D90 Keywords: online, anonymity, text data, fake news, machine learning

*We thank Sara Ekman, Therése Jonsson and Didrik Prohorenko for their excellent RA work. Furthermore, we thank Sascha Becker, Björn Tyrefors and Lucie Giorgi for fruitful comments and feedback on our work. Joakim Jansson is grateful for financial support from Jan Wallanders och Tom Hedelius stiftelse.

[†]Corresponding author: Joakim Jansson (email: joakim.jansson@lnu.se), Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden, and the Research Institute of Industrial Economics, 102 15 Stockholm, Sweden.

[‡]Department of Sociology, Uppsala University, 751 26 Uppsala, Sweden, and Swedish Institute for Social Research, Stockholm University, 106 91 Stockholm, Sweden. E-mail: emma.von.essen@soc.uu.se

1 Introduction

False information online has become a central public concern in advanced economies. In a Pew survey, 70% of respondents described it as a major threat to their country (Poushter et al. 2022).¹ A rapidly growing research literature on false information, fake news, and conspiracy theories is motivated by the concern that deteriorating information quality, especially online, can undermine public discourse, worsen policy proposals, fuel populist narratives, help elect politicians with weak platforms, or incite violence (cf. Chopra et al. 2022; Aridor et al. 2024; Meyer et al. 2024; Enders et al. 2024; Szeidl and Szucs 2025).² To date, this literature has mainly studied “soft” interventions to reduce or manage the consequences of fake-news sharing on social media, including nudges (Pennycook et al. 2020; Roozenbeek et al. 2021; Pennycook et al. 2021; Pennycook and Rand 2022; Arechar et al. 2023), sharing frictions (Fazio 2020; Henry et al. 2022), and fact-checking or digital-literacy interventions (Barrera et al. 2020; Guess et al. 2020; Roozenbeek et al. 2022; Kreps and Kriner 2022; Henry et al. 2022; Martel and Rand 2023, 2024; Meyer et al. 2024; Berger et al. 2025).³ However, relatively little evidence exists on more accountability-based policies, such as user or group bans, downvoting mechanisms, and real-name or non-anonymity requirements.⁴ Unlike nudges or information treatments, these policies do not aim to change beliefs about the accuracy of fake news. Instead, they alter the expected costs of posting false information.

This paper studies how anonymity affects fake news sharing on social media.⁵ We ask whether users reduce references to fake-news sites when they face a higher risk of being personally identified and exposed nationwide, and whether any such response is specific to fake news or also extends to legitimate news sources. A decline in fake-news sharing, with no corresponding decline in references to legitimate news sources, would suggest that accountability improves the information environment rather than merely reducing news sharing more broadly.

We answer this question using a natural experiment on Flashback, the by far largest Swedish online discussion forum. Flashback is best described as a Swedish-language counterpart to

¹In that survey, false information ranked second, behind only global climate change.

²For example, the Harvard Kennedy School has launched an academic journal fully dedicated to misinformation; see the HKS Misinformation Review, <https://misinforeview.hks.harvard.edu/>.

³There is also a large literature studying real-world consequences of various types of online content; see Aridor et al. (2024) for an overview.

⁴A recent working paper investigates the effect of wide-scale topic bans; see Di Liza and Shevchenko (2026).

⁵We follow the definition of a social media platform in Aridor et al. (2024), where social refers to content generated by users in interaction with other users, media refers to the platform operating as a two-sided market with users and advertisers, and platform denotes that these interactions take place online. Flashback, the platform we study, fulfills all three criteria.

Reddit in the breadth and scope of its discussions.⁶ Our data cover the universe of user-generated posts ($n = 1,984,224$) written by 48,672 active users in discussions related to domestic politics, immigration, and feminism between January 1, 2012, and December 31, 2016. On September 10, 2014, journalists revealed that they had obtained the identities of Flashback users who had registered before March 2007. They subsequently disclosed the identities of a small number of these early users in national media, demonstrating that anonymous usernames could be linked to offline identities and that users' online behavior could be publicized. The event created a sharp accountability shock: users registered before March 2007 faced an increased risk of losing anonymity after September 10, 2014, whereas later registrants did not.⁷ This event thus yields a canonical difference-in-differences setting that we exploit in this paper. We compare users at risk of exposure with users not at risk, before and after the event.

Flashback has several features that make it well suited for this analysis. First, it is Sweden's largest and oldest discussion forum, with more than 1.6 million registered accounts and no close substitute with comparable reach and interactivity.⁸ These features imply that established users have strong incentives to remain on the platform. Second, moderation by administrators is limited. Third, users cannot delete or edit their posts. Fourth, content is displayed chronologically rather than through an algorithmic feed. These features allow us to observe intact user-generated content and ensures that our estimates are not confounded by changes in platform algorithms or algorithmic content ranking.

Our main results show that users at risk of losing their anonymity significantly reduce the share of posts referencing fake-news sites relative to users not at risk. The estimated effect corresponds to a reduction of approximately one-quarter of the mean and is driven by a decrease in the number of posts containing links to fake-news sites, rather than by an increase in posts without fake news. We further show that the effect is concentrated in discussions related to immigration and domestic politics, with no detectable effect in discussions of feminism.⁹ The effect size can be compared to our placebo test, where we use references to the websites of Sweden's major newspapers as an outcome. We find no detectable effect on references to these legitimate news sites; the estimated coefficient corresponds to approximately 1.5% of the mean.

⁶See <https://www.flashback.org/> and <https://www.reddit.com/>.

⁷Additional details on the event are provided in Section 3 and in online appendix Section 6.2.

⁸Sweden has a population of about 10 million people. In 2018, 33% of Swedes reported having used Flashback in some form (Davidsson et al. 2018)

⁹For an analysis of how the event changed levels of xenophobia, misogyny, and overall user activity among users at risk of losing their anonymity on Flashback, see (von Essen and Jansson 2026).

The event therefore appears to reduce references to fake-news sites specifically, rather than references to news sources more generally.

There are two key concerns with our empirical strategy, in addition to the standard parallel-trends assumption which we assess using event-study graphs. First, in a canonical difference-in-difference setting, treatment of the standard errors may prove difficult (cf. Bertrand et al. 2004; Donald and Lang 2007; Conley and Taber 2011). We address this concern by collapsing the data into a time series of differences between the two groups, following Pettersson-Lidbom and Thoursie (2013), which give us correct standard.¹⁰ Second, early adopters, who faced the risk of losing their anonymity, could respond by creating new accounts. We find no evidence of such behavior. Fake-news-sharing users at risk of losing their anonymity do not exhibit an increased relative exit rate, and there is no large inflow of new accounts after the event. Moreover, our main results remain unchanged when we exclude after the event created user accounts.¹¹

The decline in fake-news sharing among users at risk of losing their anonymity could reflect three behavioral responses. Users who previously shared fake news may exit the platform; they may remain active and post at similar rates while including fake-news links less often; or they may remain active but post less frequently. Our results do not support the former two explanations, only the latter: users with a history of sharing fake news remain active, but reduce their overall posting frequency. These results suggest that the intervention does not alter users' perceptions of fake news. If it did, we would expect it to reduce their willingness to share fake news specifically, rather than their overall willingness to post.¹²

Our results can be interpreted in light of previous theoretical work. First, Glaeser (2005) argues that political hatred can be sustained through repeated false stories about out-groups.¹³ In our setting, this argument implies that fake-news references should be more common in immigration discussions and in xenophobic posts than in discussions of feminism or in misogynistic posts, since immigrants are the salient out-group in this context whereas women constitute roughly half of the Swedish population. This prediction is consistent with the patterns in our

¹⁰Specifically, it is the correct standard errors from a design-based approach to inference. In practice, we collapse the data to the weekly level, which gives us 261 observations of the difference in the weekly share of fake-news references between the two groups. We then apply the Newey–West estimator to this time series, using the number of lags suggested by the formula in Lazarus et al. (2018).

¹¹A final concern is the possibility of spillover effects on users not directly exposed to the event. However, given the continued overall upward trend in fake-news posting on the forum, this concern is unlikely to be quantitatively important. In addition, any such spillovers would likely bias our estimates toward zero.

¹²This pattern is also consistent with evidence that adding frictions to posting decreases the prevalence of fake-news content (Guriev et al. 2023).

¹³Specifically, the abstract states: “Hatred is fostered with stories of an out-group’s crimes, but the impact of these stories comes from repetition not truth” (see page 1 in Glaeser (2005)).

data. Posts classified as xenophobic are about twice as likely to contain a fake-news reference, whereas posts classified as misogynistic are no more likely than the average post to contain such a reference.¹⁴ Second, the event we study, in which a specific group of users faced a higher risk of losing their anonymity, can be interpreted as an increase in accountability for online posting behavior. This interpretation relates to the model in Acemoglu et al. (2024), which suggests that individuals seek to avoid the reputational costs of spreading misinformation, and to Szeidl and Szucs (2025), which discusses how conspiracy theories may be used by populist politicians to reduce accountability. Again, our results are consistent with these theoretical predictions, as we see users at risk of losing anonymity sharing fake-news less frequently. Moreover, this effect is particularly pronounced in immigration discussions, where xenophobic posts and fake-news references are more common, and insignificant in discussions of feminism.

Overall, our results are consistent with theoretical predictions and suggest that anonymity can facilitate the circulation of low-quality information, potentially due to limited reputational costs. They also point to the need for additional research on anonymous information provision in other settings, such as anonymous testimony in legal proceedings or the use of anonymous sources in journalism.

The remainder of the paper is organized as follows. Section 2 describes the data and provide some additional information on Flashback, Section 3 outlines the empirical strategy. Section 4 presents the main results, and Section 5 concludes.

2 Data

2.1 The online platform Flashback.org

The data in this paper stem from user-generated text messages, hereafter referred to as posts, on the Swedish web forum Flashback.org. Flashback is a large and long-standing Swedish online discussion forum, established around 2000, and is comparable to Reddit in its structure and emphasis on anonymity and free speech. Like Reddit, it hosts a wide range of discussions, including politics, sports, news media, illegal drugs, current criminal cases, and computer support. All posts are publicly accessible, and participation requires only an email address and an anonymous username. Content moderation on Flashback is very limited and mainly aims to keep

¹⁴The measures of xenophobia and misogyny come from (von Essen and Jansson 2026), which provides further details. In short, we combine hand-coded data from a research assistant with a standard BERT language model to detect xenophobic and misogynistic content.

discussions on topic.¹⁵ A 2018 survey found that 33% of Swedes had used Flashback (40% of men and 26% of women) (Davidsson et al. 2018). Flashback follows the structure of traditional message boards: discussions are organized hierarchically into subforums, threads, and posts, with posts displayed in chronological order. We focus on political discussions and therefore scraped all discussions from three subforums, Domestic politics, Immigration and integration, and Feminism, from January 1, 2012, through December 31, 2016. Overall, the collected data comprises 1,984,224 posts written by 48,672 users.

2.2 Measures and basic descriptives

Our main measure is based on hyperlinks to known fake-news websites, a measure commonly used in related research (cf. Allcott and Gentzkow 2017; Aridor et al. 2024). Two research assistants compiled the list of fake-news sites in August 2021 using external online sources from several fact-checking sites; further details are provided in Online Appendix Section 6.1. The final list comprises 2,214 hyperlinks, and is available through a Harvard Dataverse repository.¹⁶ We use this list to construct a binary indicator for whether each post contains a hyperlink to one of the identified fake-news domains. We also extract several metadata variables from each post, including the subforum name, thread title, username, timestamp, and the user’s registration date. Finally, we add indicators for whether a post contains xenophobic or misogynistic content.¹⁷

Figure 1 provides an overview of the prevalence of fake-news references. Panel 1a shows that links to fake-news sites are relatively rare. On average, approximately 1 in every 250 posts contains such a reference. By comparison, references to the websites of Sweden’s four major newspapers (Aftonbladet, Expressen, Dagens Nyheter, and Svenska Dagbladet) are about four times as common: 1 in every 67 posts contains such a link; see Figure A1 in the Online Appendix.

Figure 1a further shows that fake-news references are slightly more common in the Immigration and integration subforum. One possible explanation is suggested by Glaeser (2005), who links political hatred toward out-groups to repeated false stories. Panels 1b and 1c report the shares of posts classified as xenophobic and misogynistic, respectively. Both types of content are common in the data. Xenophobia is most frequent in immigration-related discussions,

¹⁵Ban data support this characterization: among banned users, 623 were excluded for advertisements, 280 for copyright infringement, and 421 for using multiple accounts, compared with only 36 bans for hateful posts.

¹⁶The list can be accessed at <https://doi.org/10.7910/DVN/C98Z7G>

¹⁷These measures were obtained from a fine-tuned BERT model trained and evaluated on 4,000 posts manually annotated by a research assistant. See (von Essen and Jansson 2026) for more details.

whereas misogyny is most prevalent in the subforum focused on feminism. Taken together, these patterns suggest that fake-news references are more closely associated with immigration topics and xenophobic content than with feminism-related topics and misogynistic content, in line with the argument in Glaeser (2005).

[Insert Figure 1 here]

To further characterize the relationship between xenophobia, misogyny, and fake-news references, Table 1 reports some raw relationships. Fake-news references are significantly more common in posts containing xenophobic content, but not in posts containing misogynistic content. Specifically, posts with xenophobic language are about twice as likely as the average post to contain a link to a fake-news site, whereas posts with misogynistic language if anything are less likely than the average post to contain such a link.

Column 3 of Table 1 examines whether threads whose first post contains a reference to a fake-news site receive more attention, measured by the number of responses. We find no significant relationship, with a coefficient that implies that a thread whose first post contains a fake-news link receives only about two more posts than a thread whose first post does not contain such a link. A possible conclusion from this is that fake news references are not used to garner attention or interactions, but rather for some more ideological reason. This is inline with previous research demonstrating that fake news sharing is primarily motivated by partisan polarization (Osmundsen et al. 2021) and partisan alignment (Pennycook et al. 2021) , and that outrage may fuel misinformation sharing (McLoughlin et al. 2024).¹⁸

[Insert Table 1 here]

Finally, Figure 2 shows the development of fake-news references on Flashback over the full sample period, from 2012 to 2016, on a half-year basis.¹⁹ The figure shows an upward trend in fake-news references over time, suggesting a worrying pattern that references to false information became more common on Flashback during the sample period. This pattern underscore

¹⁸There is also a literature focusing on personality traits, such as cognitive capacity and analytical thinking, and on how these traits affect the likelihood of sharing fake news (cf. Pennycook and Rand 2019; Arechar et al. 2023).

¹⁹The series is presented at the half-year level to reduce noise in the data.

the need to use our difference-in-difference strategy; to rule out this general trend in fake news sharing.

3 The event and empirical strategy

3.1 The event

On September 10, 2014, Swedish and international media revealed that a journalist collective, Researchgruppen, had obtained information that allowed them to link the usernames of Flashback users registered before March 2007 to their offline identities.²⁰ Flashback’s owner confirmed both the breach and which users were at risk of losing their anonymity in a forum thread that attracted over 300,000 views.²¹ In early 2015, Researchgruppen published several news articles exposing four users, two of whom were politicians, and stated that the goal was to discourage hateful speech online. None of the exposed users had written illegal content. Even so, the two politicians lost their positions. Since these initial disclosures, a small number of additional users, typically public figures or otherwise publicly visible individuals, have also been exposed in national media.

Google Trends data and searches of Swedish news articles containing the keywords Researchgruppen and Flashback, originally collected and presented in von Essen and Jansson (2026), show two peaks in attention: September 2014, when the breach was first revealed, and February 2015, when the exposures were published, see Figure A2. Users registered before March 2007 therefore faced a clear and verifiable risk of losing anonymity, whereas later users did not. Given the extensive debate and the administrator’s clarifications, it is reasonable to assume that users for whom exposure risk was relevant could determine whether they were at risk. Consequently, we treat users registered before March 2007 as the treated group, that is, users at risk of exposure, and later registrants as the control group, that is, users not at risk.

3.2 Empirical strategy

We use a canonical difference-in-differences (DiD) design to examine how users at risk of losing their anonymity changed their posting behavior after the shock. Our primary outcome is the

²⁰Chen, A. (2014). “The Troll Hunters.” MIT Technology Review, <https://www.technologyreview.com/s/533426/the-troll-hunters/>. In Online Appendix Section 6.2, we explain in detail how the group of journalists could connect users’ identities to their usernames and how the forum owners could verify which users risked exposure.

²¹The thread is titled “Researchgruppen påstår att de har Flashbacks databas (del 2)” and can be found at <https://www.flashback.org/t2526334p1>.

share of posts that contain references to fake-news sites.

The event described above provides one treated group, one control group, and pre- and post-treatment periods, as summarized in equation 1. Because treatment occurs at a single point in time and treatment status is fixed, recent staggered-adoption DiD estimators, such as Callaway and Sant’Anna (2021), are not required for our empirical strategy (de Chaisemartin and D’Haultfoeulle 2026).

$$Y_{tg} = \alpha + \beta \text{treated}_g \times \text{Post}_t + \gamma \text{treated}_g + \omega \text{Post}_t + \varepsilon_{tg}. \quad (1)$$

Although our setting ensures that we do not need to use staggered-adoption DiD estimators, it instead follows that estimating standard errors is less straightforward (cf. Bertrand et al. 2004; Donald and Lang 2007; Conley and Taber 2011). This follows from the fact that we have only two groups and one event. This makes a design-based approach to inference challenging, since we cannot cluster the standard errors at the group level, $g \in \{\text{Risk losing anonymity, not at risk}\}$, with only two groups. To address this concern, we construct a single time series using a two-step aggregation procedure similar to the application of Donald and Lang (2007) in Pettersson-Lidbom and Thoursie (2013). Specifically, we first collapse the data to the weekly level separately for users at risk of losing anonymity and users not at risk. This yields two group-level time series measured at weekly intervals over 261 weeks. We then express the difference in outcomes between the two groups as a single time series:

$$\bar{Y}_{t1} - \bar{Y}_{t0} = \Delta Y_t = \pi + \beta \text{Post}_t + \Delta \varepsilon_t. \quad (2)$$

Estimating Equation (2) is analytically equivalent to the canonical two-group difference-in-differences specification in Equation (1), but is implemented using aggregated weekly group-level data. In this formulation, the coefficient β is the difference-in-differences estimator. We estimate Equation (2) using the Newey–West estimator, where the key remaining choice is the number of lags to use in the estimation. Here, we follow the recommendation in Lazarus et al. (2018), which gives us an optimal number of lags equal to 21.²² This procedure allows us to estimate standard errors correctly, even though we use a canonical DiD design. We also show that we obtain similar results using a standard difference-in-differences specification equivalent

²²Specifically, the formula in Lazarus et al. (2018) states that the number of lags should equal $S = 1.3T^{1/2}$, where S is the number of lags and T is the number of time periods. Inserting $T = 261$ gives $S = 21$.

to Equation (1). In the Online Appendix, we provide estimates using fewer or more lags in the Newey–West estimator.

4 Results

4.1 Main results

We start by presenting standard event-study figures at the year and half-year levels in Figures 3a and 3b.²³ The pre-event coefficients are close to zero and follow a flat trend, thus demonstrating that the upward aggregate trend in fake-news references shown in Figure 2 is resolved by our difference-in-difference methodology. Further, parallel trends in the absence of treatment seem to be a reasonable assumption based on these figures. After the event, the coefficient estimates decline sharply, implying that users at risk of losing their anonymity reduce references to fake-news sites relative to users not at risk.

[Insert Figure 2 here]

Next, Table 2 presents our main regression results from the standard and collapsed difference-in-differences models in Panels A and B, respectively. In column 1, the DiD estimate shows that users at risk of losing their anonymity reduce the share of posts containing references to fake-news websites after the event. The estimated decrease is 0.1 percentage points, which corresponds to about one-quarter of the mean of 0.004 (or 0.4 percentage points) shown in Figure 1a. The effect is thus relatively large. Columns 2–4 report results by subforum. The estimates are similar in magnitude in the Domestic politics and Immigration and integration subforums, but we find no statistically significant decrease in the Feminism subforum.

[Insert Table 2 here]

We use references to legitimate news sources as a placebo outcome in our baseline difference-in-differences model. As noted above, we interpret the risk of losing anonymity as a shock to accountability. Unlike sharing fake-news, sharing ordinary news sources should not expose users

²³The standard errors used to construct the confidence intervals in these figures are estimated by clustering at the thread level. Thus, they are formally not correct. However, as Table 2 shows, they are similar in magnitude to the standard errors from our main specification.

to reputational costs. We therefore do not expect the event to reduce references to legitimate news sources, making this an informative placebo test of our empirical strategy. As discussed in Section 2.2, we measure legitimate news sources using links to the websites of Sweden’s four major newspapers. The results, reported in Panels C and D of Table 2, show no statistically significant effect on sharing legitimate news sources, with a coefficient corresponding to 1.5% of the mean.

4.2 Mechanisms

Ex ante, several behavioral changes could explain the decline in the share of posts containing fake-news links. First, the observed decrease could, in principle, be driven either by fewer posts containing links to fake-news sites or by more posts without such links. We show that the effect is driven by a decrease in the number of posts containing links to fake-news sites; see Online Appendix Table A1.

Following this, the decrease could arise due to three channels among users at risk of losing their anonymity. First, users who share fake news could abandon the forum altogether. Second, they could remain active but write fewer posts and contribute less content. Third, they could remain active and write as many posts as before, but include links to fake-news sites less often. In Online Appendix Tables A2–A4, we distinguish between these three potential channels. Table A2 provides evidence against the first channel. Specifically, users at risk of losing their anonymity who had referenced fake news before the event are no more likely to stop posting than users at risk of losing their anonymity without a history of fake-news references. This suggests that the effect is not driven by these users abandoning the platform, but rather by changes in their posting behavior while they remain active contributors.

Next, to distinguish between the second and third channels, less frequent posting or changes in the content of posts, we study users who write at least one post before and one post after the event in a balanced two-period panel.²⁴ To test whether the content of posts changes after the event, we examine whether the individual-level share of posts containing fake-news links decreases in Online Appendix Table A3. We find no change in this share in the balanced individual-level panel. This suggests that the main results are not driven by active users selectively removing fake-news links from otherwise similar posting activity. Instead, the results

²⁴Data is thus structured at the individual-user level, with one observation before the event and one after. Each observation contains the user’s average share of posts containing fake-news links and the number of posts written by the user during that time period.

point to users with a history of sharing fake news remaining active on the platform but writing fewer posts overall, which is exactly what we find in Online Appendix Table A4.

4.3 Robustness tests

A key threat to our empirical strategy is if users after the event created new accounts to continue posting links to fake-news sites. This concern would be especially egregious if accounts registered after the event drive the results by accounting for a disproportionate share of fake-news references. We therefore exclude all accounts registered after the event in Table A5. The results are unchanged, which is consistent with the finding above that users at risk of exposure with a history of sharing fake news are no more likely to stop posting. In Online Appendix Figure A3, we also examine users' registration month in our data and find no large inflow of accounts registered after the event, as indicated by the solid red line in the figure.²⁵

Another empirical question is whether 21 lags in the Newey–West estimator adequately capture the degree of auto-correlation of the standard errors in our time series. We therefore replicate Panel B of Table 2 using 8, 16, and 32 lags in the Newey–West estimator; see Online Appendix Tables A6, A7, and A8, respectively. These alternatives do not change the estimated standard errors or our main results to any meaningful degree.

5 Concluding discussion

In this paper, we show how the risk of losing anonymity affects references to fake-news sites on social media. In line with predictions from previous theoretical work (see Glaeser 2005; Acemoglu et al. 2024; Szeidl and Szucs 2025), we find that users who face a higher risk of losing anonymity reduce the share of posts containing fake-news references. This effect is concentrated in discussions of immigration, but is not statistically detectable in discussions of feminism. Conversely, we find no decrease in how often users reference standard news outlets, suggesting that the event improved the information environment rather than reducing news sharing more broadly. This distinction is important because policies aimed at reducing fake-news sharing may also backfire by reducing information sharing overall, as demonstrated in Ershov and Morales (2024).

²⁵Another related concern is the possibility of spillover effects on users not directly exposed to the event. However, given the continued overall upward trend in fake-news posting on the forum, this concern is unlikely to be quantitatively important. Moreover, any such spillovers would likely bias our estimates toward zero.

The effect size we estimate is about one-quarter of the mean. A key question is how this magnitude compares with the “softer” interventions studied in previous research. Before comparing effect sizes, it is important to note that most previous studies have been conducted on online crowd-sourcing platforms such as MTurk. Few papers study real-world applications, which may affect estimated effect sizes. With this caveat in mind, the nudging literature typically finds smaller effects, ranging from 3.5% to 20% of the control-group mean (Pennycook et al. 2020; Roozenbeek et al. 2021; Pennycook et al. 2021; Arechar et al. 2023). Sharing frictions, though less commonly studied, have estimated effects of around 18% to 32% of the control-group mean (Fazio 2020). The fake-news labeling literature includes two meta-analyses: one reports effect sizes between 25% and 45% of the mean (Martel and Rand 2023), while the other reports a 25% decrease in sharing. Thus, our estimated effect is toward the upper end of the range reported in the literature. Moreover, one potential advantage of increasing accountability is that it may be relatively straightforward to implement in some online settings. Compared with labeling fake-news posts, for instance, it is less labor-intensive and avoids disputes over which content should be labeled as false information.

Still, although our results suggest an improved information environment, we do not wish to advocate non-anonymity across all platforms or contexts. Accountability appears important for reducing references to false information. However, increasing accountability may also have costs, such as discouraging users from engaging in certain discussions, topics, or arguments. Thus, while our paper is, to our knowledge, the first to study the relationship between anonymity, accountability, and misinformation sharing, it should not be the last. Future research should examine other settings in which anonymity may affect the reliability of information. Anonymous witnesses in court, anonymous referees in peer-reviewed research, and anonymous sources in journalism are examples of settings in which our paper points to the need for empirical evidence on how anonymity affects the quality of the information provided. Similarly, our paper highlights the need for further research on how anonymity affects other dimensions of online content, such as hate speech, xenophobia, misogyny, topic avoidance, and political expression. There is also a need to assess other commonly used accountability-based policies, such as bans, downvoting, and content moderation, and how they affect the sharing of fake news on social media.

References

- Acemoglu, D., A. Ozdaglar, and J. Siderius (2024). A model of online misinformation. *Review of Economic Studies* 91(6), 3117–3150.
- Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2), 211–36.
- Arechar, A. A., J. Allen, A. J. Berinsky, R. Cole, Z. Epstein, K. Garimella, A. Gully, J. G. Lu, R. M. Ross, M. N. Stagnaro, et al. (2023). Understanding and combatting misinformation across 16 countries on six continents. *Nature Human Behaviour* 7(9), 1502–1513.
- Aridor, G., R. Jiménez-Durán, R. Levy, and L. Song (2024). The economics of social media. *Journal of Economic Literature* 62(4), 1422–1474.
- Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of public economics* 182, 104123.
- Berger, L. M., A. Kerkhof, F. Mindl, and J. Münster (2025). Debunking “fake news” on social media: Immediate and short-term effects of fact-checking and media literacy interventions. *Journal of Public Economics* 245, 105345.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics* 119(1), 249–275.
- Callaway, B. and P. H. Sant’Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics* 225(2), 200–230.
- Chopra, F., I. Haaland, and C. Roth (2022). Do people demand fact-checked news? evidence from us democrats. *Journal of Public Economics* 205, 104549.
- Conley, T. G. and C. R. Taber (2011). Inference with “difference in differences” with a small number of policy changes. *The Review of Economics and Statistics* 93(1), 113–125.
- Davidsson, P., M. Palm, and A. Melin Mandre (2018). *Svenskarna och Internet 2018*. IIS (Internetstiftelsen i Sverige).
- de Chaisemartin, C. and X. D’Haultfoeulle (2026). Credible answers to hard questions: Differences-in-differences for natural experiments. Available at SSRN 4487202.

- Di Liza, A. and L. Shevchenko (2026). Can hate speech be banned online? the effects of shutting down toxic forums on reddit.
- Donald, S. G. and K. Lang (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics* 89(2), 221–233.
- Enders, A., C. Klostad, and J. Uscinski (2024). The relationship between conspiracy theory beliefs and political violence. *Harvard Kennedy School Misinformation Review* 5(6).
- Ershov, D. and J. S. Morales (2024). Sharing news left and right: Frictions and misinformation on twitter. *The Economic Journal* 134(662), 2391–2417.
- Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review* 1(2).
- Glaeser, E. L. (2005). The political economy of hatred. *The Quarterly Journal of Economics* 120(1), 45–86.
- Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar (2020). A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences* 117(27), 15536–15545.
- Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). Curtailing false news, amplifying truth. *Amplifying Truth* (October 29, 2023).
- Henry, E., E. Zhuravskaya, and S. Guriev (2022). Checking and sharing alt-facts. *American Economic Journal: Economic Policy* 14(3), 55–86.
- Kreps, S. E. and D. L. Kriner (2022). The covid-19 infodemic and the efficacy of interventions intended to reduce misinformation. *Public Opinion Quarterly* 86(1), 162–175.
- Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018). Har inference: Recommendations for practice. *Journal of Business & Economic Statistics* 36(4), 541–559.
- Martel, C. and D. G. Rand (2023). Misinformation warning labels are widely effective: A review of warning effects and their moderating features. *Current Opinion in psychology* 54, 101710.
- Martel, C. and D. G. Rand (2024). Fact-checker warning labels are effective even for those who distrust fact-checkers. *Nature Human Behaviour* 8(10), 1957–1967.

- McLoughlin, K. L., W. J. Brady, A. Goolsbee, B. Kaiser, K. Klonick, and M. Crockett (2024). Misinformation exploits outrage to spread online. *Science* 386(6725), 991–996.
- Meyer, M., A. Enders, C. Klostad, J. Stoler, and J. Uscinski (2024). Using an ai-powered “street epistemologist” chatbot and reflection tasks to diminish conspiracy theory beliefs. *Harvard Kennedy School Misinformation Review* 5(6).
- Osmundsen, M., A. Bor, P. B. Vahlstrup, A. Bechmann, and M. B. Petersen (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., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand (2021). Shifting attention to accuracy can reduce misinformation online. *Nature* 592(7855), 590–595.
- Pennycook, G., J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand (2020). Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science* 31(7), 770–780.
- Pennycook, G. and D. G. Rand (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188, 39–50.
- Pennycook, G. and D. G. Rand (2022). Nudging social media toward accuracy. *The Annals of the American Academy of Political and Social Science* 700(1), 152–164.
- Pettersson-Lidbom, P. and P. S. Thoursie (2013). Temporary disability insurance and labor supply: evidence from a natural experiment. *The Scandinavian Journal of Economics* 115(2), 485–507.
- Poushter, J., M. Fagan, and S. Gubbala (2022). Climate change remains top global threat across 19-country survey. Pew Research Center.
- Roozenbeek, J., A. L. Freeman, and S. Van Der Linden (2021). How accurate are accuracy-nudge interventions? a preregistered direct replication of pennycook et al.(2020). *Psychological science* 32(7), 1169–1178.
- Roozenbeek, J., S. Van Der Linden, B. Goldberg, S. Rathje, and S. Lewandowsky (2022). Psychological inoculation improves resilience against misinformation on social media. *Science advances* 8(34), eabo6254.

Szeidl, A. and F. Szucs (2025). A model of populism as a conspiracy theory. *American Economic Review* 115(9), 3214–3247.

von Essen, E. and J. Jansson (2026). Misogynistic and xenophobic hate language online: a matter of anonymity.

Table 1: Simple correlations

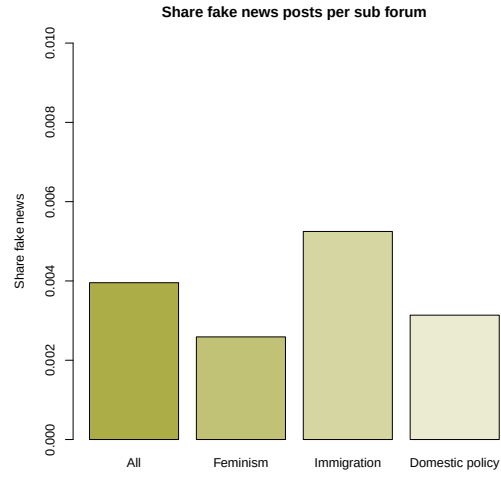
	Dependent variable:		
	P(Fake news)		Thread length
	(1)	(2)	(3)
Xenophobic post	0.00415**** (0.00018)		
Misogynistic post		-0.00030* (0.00017)	
1st entry fake news			2.20383 (6.41878)
Observations	1984224	1984224	29455

Note: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.
The standard errors are clustered at the thread level (29 455 clusters). The dependent variable is a dummy indicating if a post contains a link to a fake news site in columns 1 and 2 and thread length in column 3. An observations is a post in columns 1 and 2 and a thread in column 3.

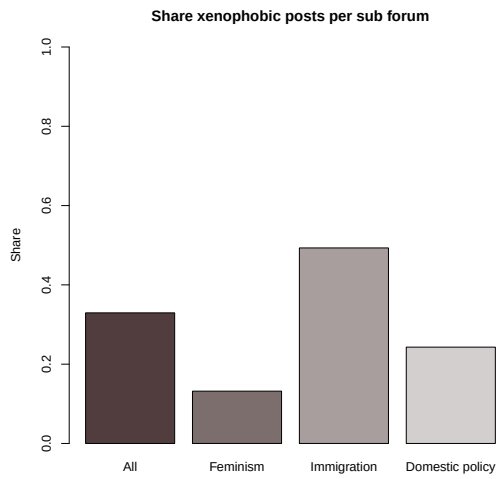
Table 2: DD estimates of a decrease in anonymity on fake news and legitimate news

	(1)	(2)	(3)	(4)
Panel A: Main outcome, standard errors clustered at the thread level				
	Dependent variable: P(Fake news)			
Risk losing anonymity	-0.00110*** (0.00034)	-0.00089** (0.00040)	-0.00112* (0.00058)	0.00003 (0.00093)
Observations	1984224	639768	887928	456528
Panel B: Main outcome, Newey-West std errs with optimal (21) lags				
	Dependent variable: P(Fake news)			
Risk losing anonymity	-0.00115*** (0.00042)	-0.00130*** (0.00042)	-0.00145** (0.00060)	-0.00038 (0.00074)
Observations	261	261	261	261
Panel C: Placebo test, standard errors clustered at the thread level				
	Dependent variable: P(Legitimate news)			
Risk losing anonymity	-0.00061 (0.00088)	-0.00055 (0.00233)	-0.00115 (0.00102)	0.00203 (0.00147)
Observations	1984224	639768	887928	456528
Panel D: Placebo test, Newey-West std errs with optimal (21) lags				
	Dependent variable: P(Legitimate news)			
Risk losing anonymity	-0.00023 (0.00138)	-0.00141 (0.00295)	-0.00057 (0.00104)	0.00159 (0.00159)
Observations	261	261	261	261
Forum	All	Domestic politics	Immigration	Feminism

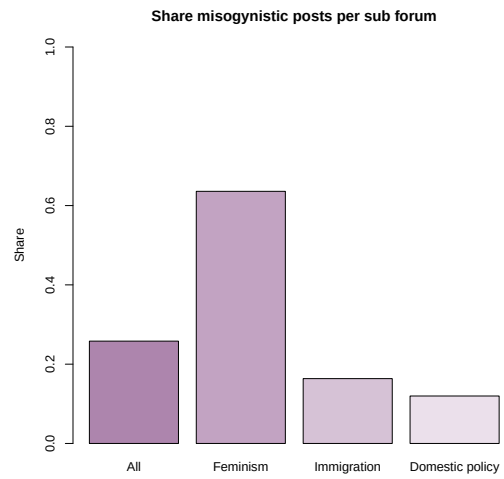
Note: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. The time period used in the regressions is from January 1st 2012 until December 31st 2016. The standard errors are clustered at the thread level (29 455 clusters) in panels A and C, and Newey-West standard errors with 21 lags on a collapsed time series in panels B and D. The dependent variable is a dummy indicating if a post contains a link to a fake news site in panels A and B, and a dummy indicating if a post contains a link to a legitimate news site in panels C and D. In panels A and C, an observation is a post at Flashback. In panels B and D data is structured as a time series of the difference between the early and late adopters at the weekly level.



(a)



(b)



(c)

Figure 1: Share of posts with fake news references, xenophobia and misogyny across subforums

The Figure presents the share of posts on Flashback containing a reference to fake news articles, xenophobia and misogyny for each separate subforum and in total in our data.

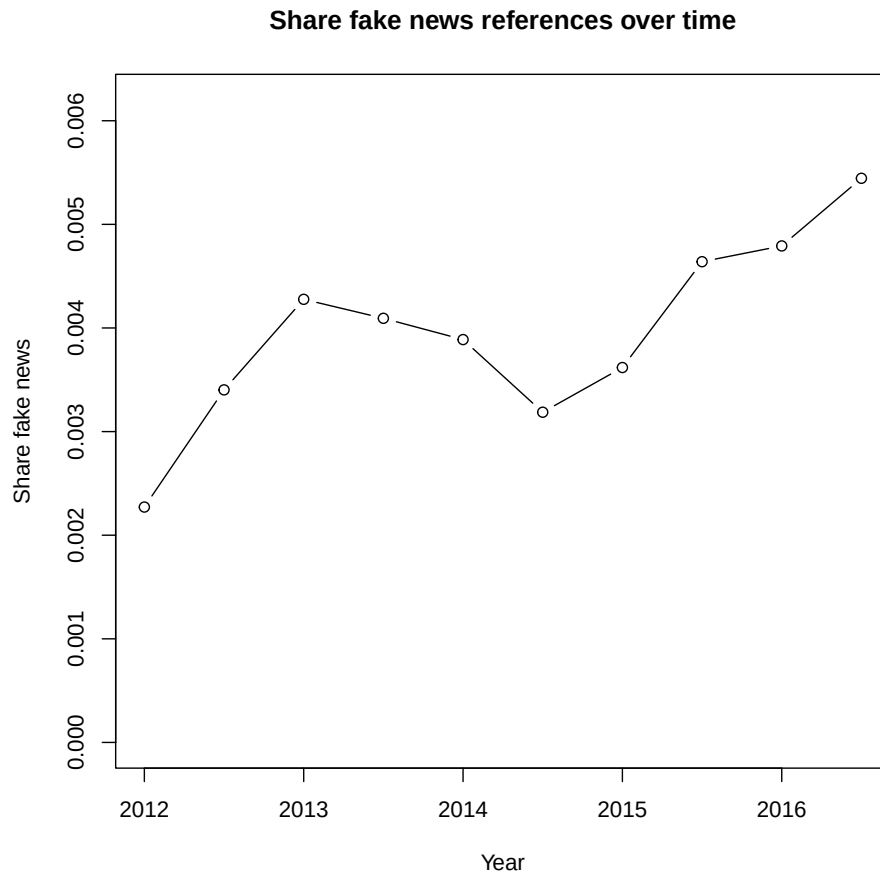


Figure 2: Trend in fake news

The Figure presents the general trend of fake news sharing on Flashback over time in the data, aggregated to the half-year level.

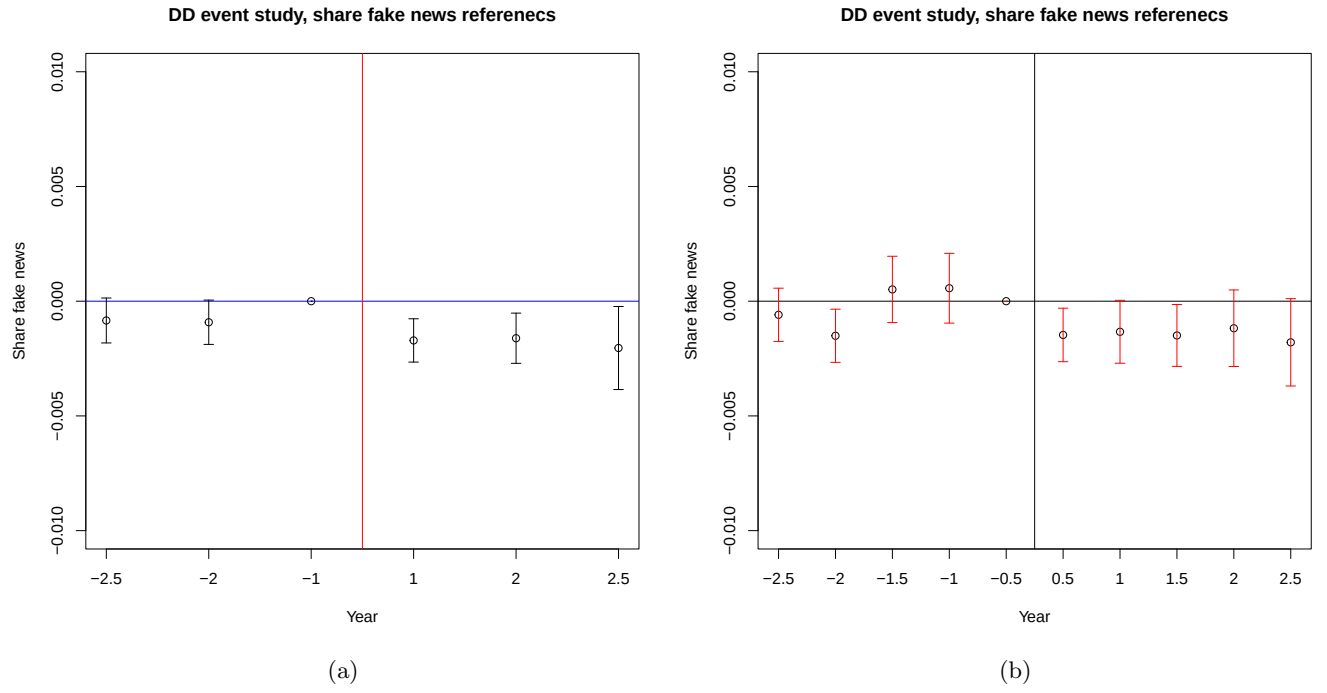


Figure 3: DD Event study - share fake news references

Panel A and B demonstrate a standard event study plot, plotting the difference between the users risking losing their anonymity and those not at risk at the yearly and half-yearly level respectively. The solid red line present in all of the panels denotes September 10 2014; the point after which the users were at risk of losing their anonymity. Confidence intervals constructed by clustering the standard errors at the thread level (32,142 clusters).

6 Online appendix

6.1 Collecting links to fake news sites

The data used in this study consist of text-based messages, referred to as posts, collected from the online Swedish discussion forum Flashback. Using custom-built Python scripts, we scraped all posts from three subforums—feminism, domestic politics, and immigration and integration—from the inception of each forum until January 2017. In von Essen and Jansson (2026) we use this data to investigate the effect of decreased anonymity on the frequency of xenophobic and misogynistic expressions. In this paper, we instead focus on the effect on using fake news sites as references.

To obtain a measure of the amount of references to fake news sites, we gave two research assistants the task of collecting and organizing links to sites deemed as fake news by various journalistic and fact checking sources. We provided the list of sources, and the RAs task was to collect and organize all the links. The full list of sources is provided among the references (Columbia Journalism Review 2021; FactCheck 2021; The Fake News Codex 2021; Daily Dot 2021; Media Bias Fact Check 1 2021; Media Bias Fact Check 2 2021; Viralgranskaren 2021; CBS News 2021; Zimdars 2021; Wikipedia 2021; PolitiFacts 2021; Snope’s 2021). The actual list of web-addresses to fake news sites is available as a separate data set at <https://doi.org/10.7910/DVN/C98Z7G>. In total, it consists of 2214 fake news sites.

With this list of fake news sites in hand, we next searched each post at Flashback to determine if it contained a link to any of the sites in our list. We generated a simple dummy measure in this instance; if a post contained a link from our list, then the dummy was assigned the value 1, and zero otherwise. Code generating this measure is naturally part of the replication package associated with this paper.

As a basic measure of referencing to actual news sites, we also checked if each entry contained a link to one of the four major news papers in Sweden: Dagens Nyheter, Svenska Dagbladet, Aftonbladet and Expressen.

6.2 Behind the event: How could the journalists access the identity of the users and how did the owner verify those at risk?

A common question that arises is how the journalists could obtain the identities for the early registered users when Flashback was supposed to be securely anonymous. In this subsection we thus outline the process that the journalists used to obtain the identities behind the early registered accounts at Flashback.

The group of journalists originally had obtained a list, a so called dump, of e-mail addresses that were connected to usernames from Flashback. The list was however from a few years back, which is why only users registered before March 2007 were in it. To this list, the users' identities were added by matching the e-mail addresses from Flashback to e-mail addresses from other leaked Internet sites, such as CDON.com, from which other data, such as Swedish social security numbers and names, could be collected. Thus, the journalists could link usernames to identities of individuals with accounts registered at Flashback before March 2007.

The journalists had a history of publicly exposing individuals writing anonymous xenophobic comments online, making the threat of future exposures credible. A spreadsheet-file containing the email addresses and identities that the journalists had access to was subsequently posted publicly online, but without the associated usernames from Flashback. The owner of Flashback could however see that the rows of the file were sorted alphabetically based on the usernames, implying that the journalists had access to the usernames as well. Furthermore, the owner could see that the accounts that were in the file were registered before March 2007. The owner of Flashback therefore declared in a thread²⁶ that those who had registered before March 2007 were at risk of being exposed. Said thread have to this day been read more than 300,000 times, thus making it clear that many users wanted to know if they were at risk. Several users expressed their disappointment and outrage that this had happened in the thread, and declared that they may leave Flashback as a consequence.

²⁶See "Researchgruppen påstår att de har Flashbacks databas (del 2)", can be found at <https://www.flashback.org/t2526334p1>.

6.3 Additional results - tables and figures

6.3.1 Figures

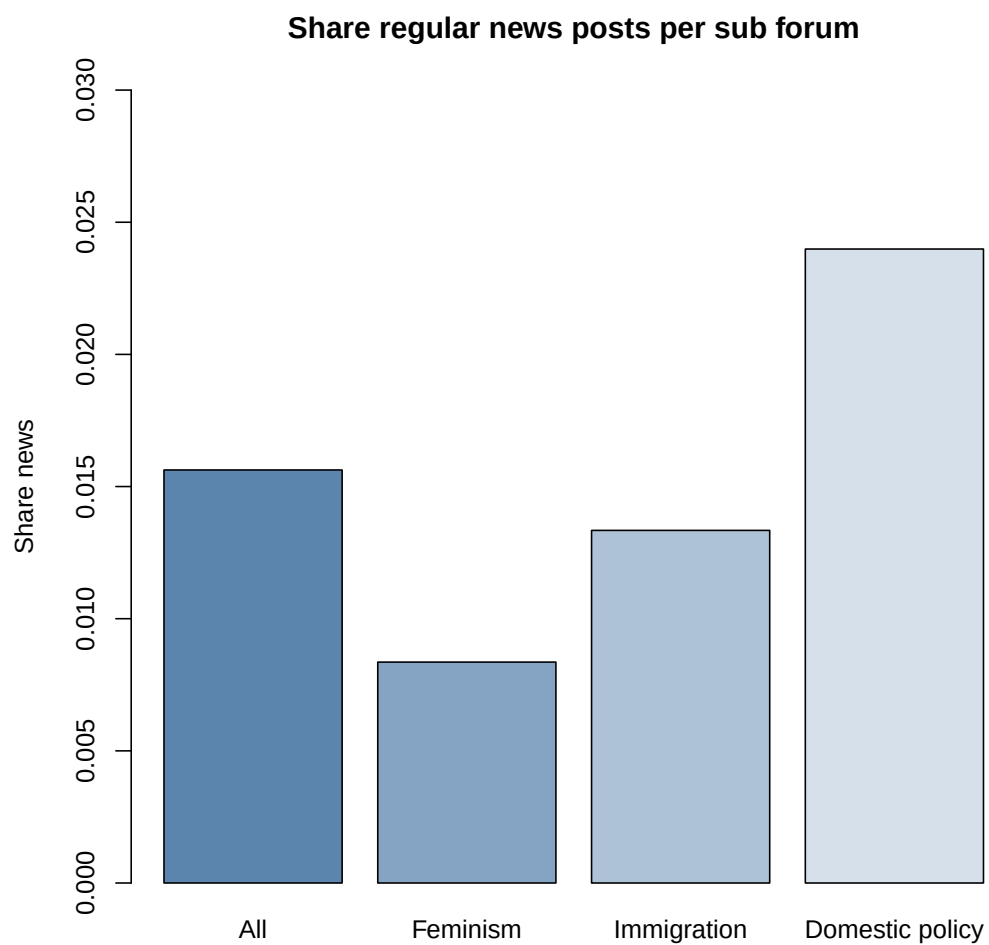


Figure A1: Share major news paper references across subforums

The figure presents the share of posts containing a reference to one of Sweden's four major news papers websites for each separate subforum and in total in our data.

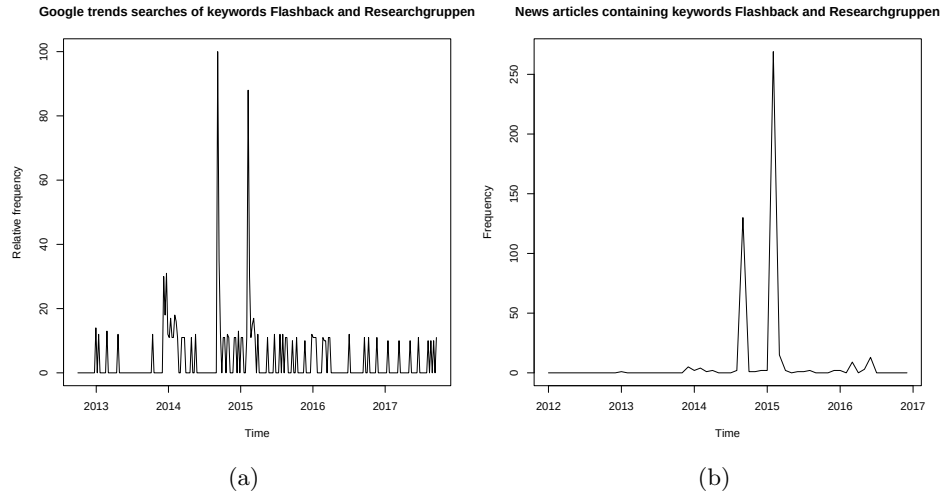


Figure A2: Google searches and news articles relating to the event

The left graph displays data from Google Trends for the weekly relative search frequencies of the terms Flashback and Researchgruppen between 2013 and 2017. The relative search frequency shows a spike if there are at least 50 searches, and all of the spikes are relative to the largest spike, which reaches 100 in the figure. Thus, a spike reaching 80 implies that the search terms reach 80% of the highest search frequency. The right graph displays a graph similar to the monthly number of news articles in Sweden that contain the words “Flashback” and “Researchgruppen”. The data on the number of news articles were obtained through searches in the Mediearkivet database (<https://www.retriever.se/>, accessed using a login from Stockholm University on July 4, 2018).

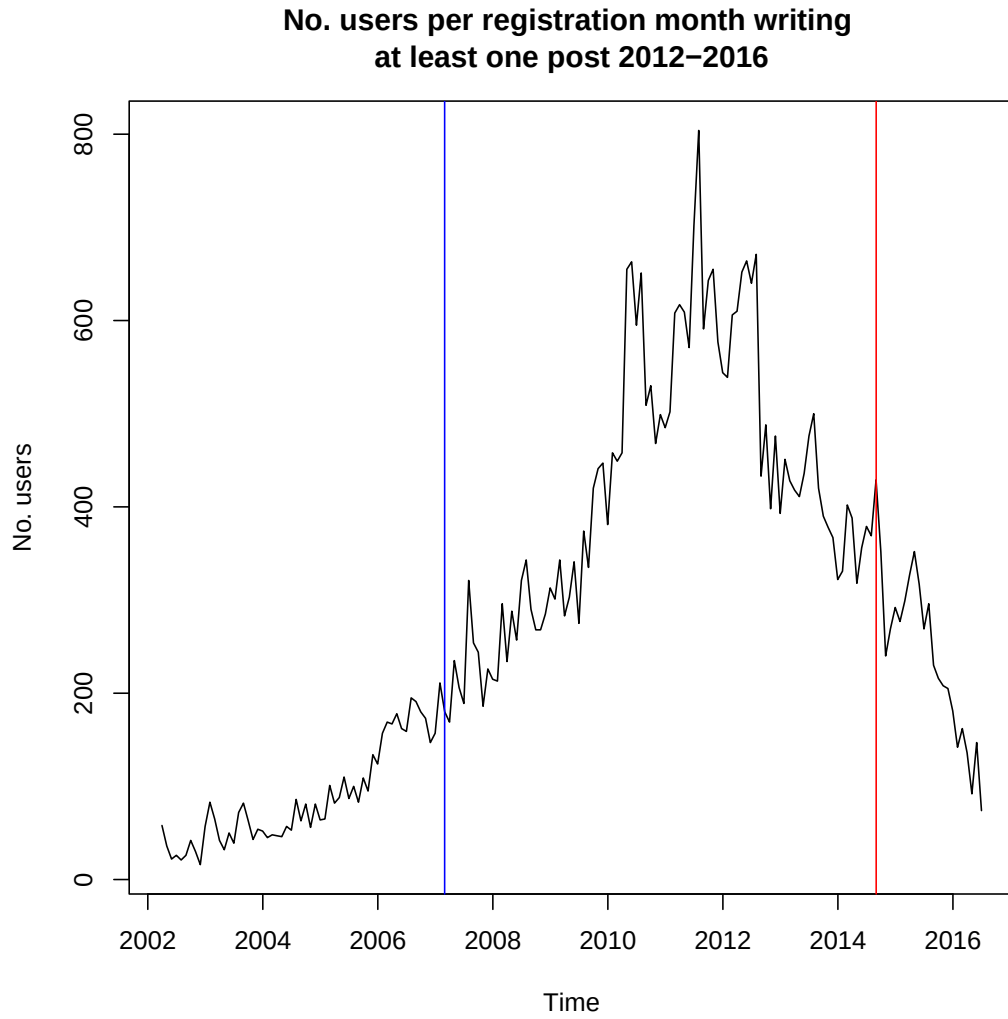


Figure A3: Users per registration month in data

The figure presents the number of users per registration month who have written at least one post in our data from 2012 to 2016. The red line lies on September 2014 and thus represents where the period after the event starts. The blue line lies on March 2007. The users below this cut-off registration date are the ones that risked exposure after September 10 2014.

6.3.2 Tables

Table A1: DD estimate of a decrease in anonymity on number of fake news posts

	Dependent variable:			
	No. fake news posts			
	(1)	(2)	(3)	(4)
Risk losing anonymity	−9.80153** (3.87909)	−2.11871** (0.81812)	−6.22509* (3.49745)	−1.45773** (0.56868)
Forum	All	Domestic politics	Immigration	Feminism
Observations	261	261	261	261

Note:

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

The time period used in the regressions is from January 1st 2012 until December 31st 2016. The standard errors are Newey-West standard errors with 21 lags on a collapsed time series. The dependent variable is the number of weekly posts containing a link to a fake news site. Data is structured as a time series of the difference between the early and late adopters at the weekly level.

Table A2: Probability of leaving the platform, heterogeneity

	Dependent variable:			
	P(Leave forum)			
	(1)	(2)	(3)	(4)
Risk losing*prior fake news ref	-0.029 (0.040)	-0.028 (0.040)	-0.028 (0.040)	-0.026 (0.040)
Risk losing anonymity	-0.148**** (0.008)	-0.052**** (0.012)	0.049*** (0.015)	0.099**** (0.016)
Prior fake news ref	-0.256**** (0.014)	-0.257**** (0.014)	-0.257**** (0.014)	-0.258**** (0.014)
Startdate control function	-	Linear	Quadratic	Cube
Observations	34460	34460	34460	34460
Observations	34,460	34,460	34,460	34,460

Note:

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

The table demonstrate the probability of a user that risk having their anonymity revoked leaving Flashback after the journalists gained access to their identities on September 10 2014. The standard errors are clustered at the user level (48 672 clusters). The dependent variable is a dummy indicating if a user leaves the platform.

Table A3: DD estimate of a decrease in anonymity on fake news, balanced panel

	Dependent variable:
	P(Fake news)
Risk losing anonymity	−0.00024 (0.00114)
Observations	22318

Note:

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

The time period used in the regressions is from January 1st 2012 until December 31st 2016.

The standard errors are clustered at the user level. Data is structured as a balanced two period panel at the user level The dependent variable is the average links to fake news sites per time period per user. An observation is a users posts pre or post event.

Table A4: DD estimate of a decrease in anonymity on number of written posts, balanced panel

	Dependent variable:	
	No. posts	
	(1)	(2)
Risk losing anonymity	-4.67989 (3.23099)	-133.92320** (63.18624)
Fake news posting history	No	Yes
Observations	20884	1434

Note:

Significance levels: *
 $p < 0.10$, ** $p < 0.05$, ***
 $p < 0.01$, **** $p < 0.001$.
The time period used in the
regressions is from January
1st 2012 until December
31st 2016. The standard
errors are clustered at the
user level. Data is
structured as a balanced
two period panel at the user
level The dependent
variable is the number
written posts per time
period per user. An
observation is a users posts
pre or post event.

Table A5: DD estimate of a decrease in anonymity on fake news, only pre event users

	Dependent variable:			
	P(Fake news)			
	(1)	(2)	(3)	(4)
Risk losing anonymity	−0.00102** (0.00042)	−0.00136*** (0.00044)	−0.00143** (0.00064)	0.00025 (0.00074)
Forum	All	Domestic politics	Immigration	Feminism
Observations	261	261	261	261

Note: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.
The time period used in the regressions is from January 1st 2012 until December 31st 2016. The standard errors are Newey-West standad errors with 21 lags on a collapsed time series. The dependent variable is a dummy indicating if a post contains a link to a fake news site. Data is structured as a time series of the difference between the early and late adopters at the weekly level.

Table A6: DD estimate of a decrease in anonymity on fake news, using 8 lags

	Dependent variable:			
	P(Fake news)			
	(1)	(2)	(3)	(4)
Risk losing anonymity	−0.00115*** (0.00039)	−0.00130*** (0.00042)	−0.00145** (0.00065)	−0.00038 (0.00089)
Forum	All	Domestic politics	Immigration	Feminism
Observations	261	261	261	261

Note:

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

The time period used in the regressions is from January 1st 2012 until December 31st 2016. The standard errors are Newey-West standard errors with 8 lags on a collapsed time series. The dependent variable is a dummy indicating if a post contains a link to a fake news site. Data is structured as a time series of the difference between the early and late adopters at the weekly level.

Table A7: DD estimate of a decrease in anonymity on fake news, using 16 lags

	Dependent variable:			
	P(Fake news)			
	(1)	(2)	(3)	(4)
Risk losing anonymity	−0.00115*** (0.00042)	−0.00130*** (0.00041)	−0.00145** (0.00061)	−0.00038 (0.00079)
Forum	All	Domestic politics	Immigration	Feminism
Observations	261	261	261	261

Note:

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

The time period used in the regressions is from January 1st 2012 until December 31st 2016. The standard errors are Newey-West standard errors with 16 lags on a collapsed time series. The dependent variable is a dummy indicating if a post contains a link to a fake news site. Data is structured as a time series of the difference between the early and late adopters at the weekly level.

Table A8: DD estimate of a decrease in anonymity on fake news, using 32 lags

	Dependent variable:			
	P(Fake news)			
	(1)	(2)	(3)	(4)
Risk losing anonymity	−0.00115*** (0.00043)	−0.00130*** (0.00042)	−0.00145** (0.00059)	−0.00038 (0.00070)
Forum	All	Domestic politics	Immigration	Feminism
Observations	261	261	261	261

Note:

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

The time period used in the regressions is from January 1st 2012 until December 31st 2016. The standard errors are Newey-West standard errors with 32 lags on a collapsed time series. The dependent variable is a dummy indicating if a post contains a link to a fake news site. Data is structured as a time series of the difference between the early and late adopters at the weekly level.

6.4 Online appendix references

References

- CBS News. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://www.cbsnews.com/pictures/dont-get-fooled-by-these-fake-news-sites/>.
- Columbia Journalism Review. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://web.archive.org/web/20210720140548/https://www.cjr.org/fake-beta>.
- Daily Dot. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://www.dailydot.com/debug/fake-news-sites-list-facebook/>.
- von Essen, Emma and Joakim Jansson. 2026. "Misogynistic and xenophobic hate language online: a matter of anonymity." IFN Working Paper.
- FactCheck. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/>.
- Media Bias Fact Check 1. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://mediabiasfactcheck.com/fake-news/>.
- Media Bias Fact Check 2. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://mediabiasfactcheck.com/conspiracy/>.
- PolitiFacts. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://infogram.com/politifacts-fake-news-almanac-1gew2vjdxl912nj>.
- Snope's. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://www.snopes.com/news/2016/01/14/fake-news-sites/>.
- The Fake News Codex. 2021. "Links to Fake News Sites." Access 2021-07-01, <http://www.fakenewscodex.com>.
- Viralgranskaren. 2021. "Links to Fake News Sites." Access 2021-07-01, <https://web.archive.org/web/20190804194427/https://www.metro.se/viralgranskaren/varningslistan-ZXUdYc8rU>.
- Wikipedia. 2021. "Links to Fake News Sites." Access 2021-07-01, https://en.wikipedia.org/wiki/List_of_fake_news_websites.

Zimdars. 2021. "Communication and Media at Merrimack College in North Andover List of Fake News." Access 2021-07-01, https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QG5ewC3VAL6pLkT53V_81ZyitM/preview.