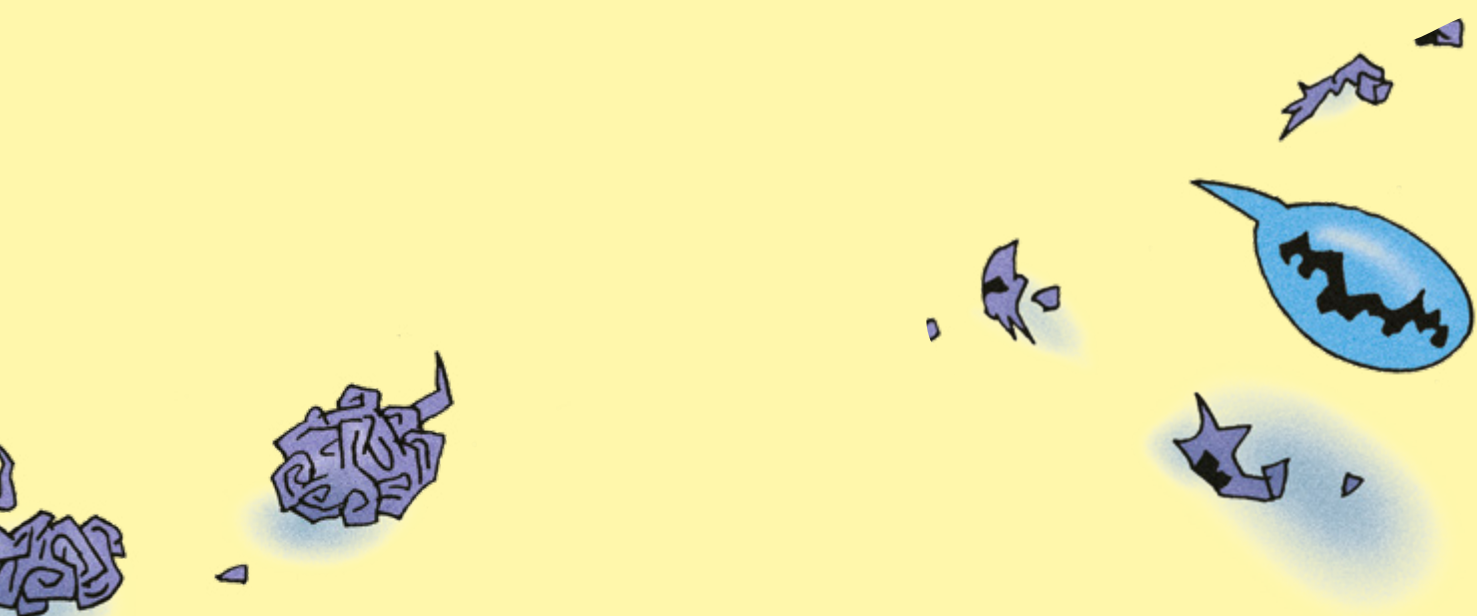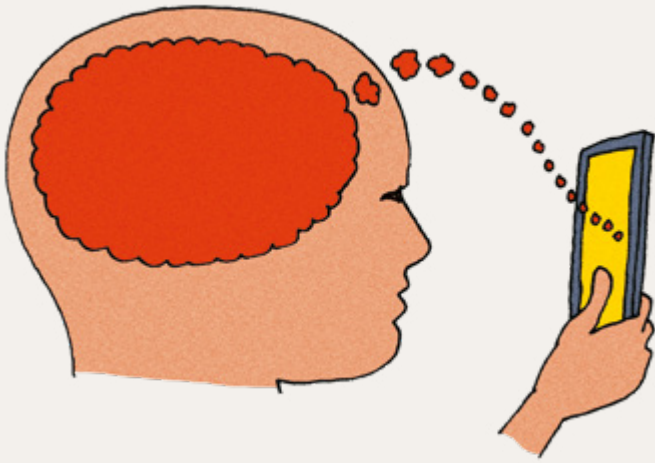Émeric Henry is a microeconomist and Director of the Département d'Économie (Department of Economics). His research focuses on the economics of innovation, the digital economy and political economy. He combines economic theory with experimental and empirical methods.

Combating misinformation on social media

By Émeric Henry

In addition to regulation and long-term policies, an inexpensive way of curbing the spread of false information online would be to take action as early and as upstream as possible, influencing internet users themselves. The desire of individuals not to appear ill-informed in the eyes of their audience, thereby damaging their reputation, could be an effective lever, as shown by the different treatments tested with a group of internet users in a recent empirical survey to which Émeric Henry contributed.

A delicate balance must be struck between combating false information and protecting freedom of expression.

Social media has fundamentally changed the way we interact, communicate and access information. Its potential to spread misinformation is a major concern for citizens and politicians alike. Political misinformation is rife on platforms such as Facebook, X/Twitter and Reddit. This is worrying given that a substantial share of users rely on these platforms to get information.

A delicate balance needs to be struck between combating false information and protecting freedom of expression. In the United States, constitutional limits hinder the regulation of content moderation. The European Union does plan to regulate platforms via the Digital Services Act (DSA), but for the time being the focus is on illegal content while significant political misinformation continues. Some researchers are advocating for the introduction of digital education programmes to teach citizens to distinguish between accurate information and fake news as a long-term solution to combat the phenomenon.

A completely different approach consists of influencing users before they decide whether or not to share content on social media, that is, taking action as early as possible. Such a policy would be less costly and some of its components would be easy to implement. It could involve requiring confirmation clicks when the decision is made to share, encouraging users to think about the consequences of sharing false information — an intervention known as a 'nudge' that was recently demonstrated to be effective by psychologist Gordon Pennycook and David Rand, professor of management science, brain and cognitive sciences — or even offering fact-checking, as some platforms already do.
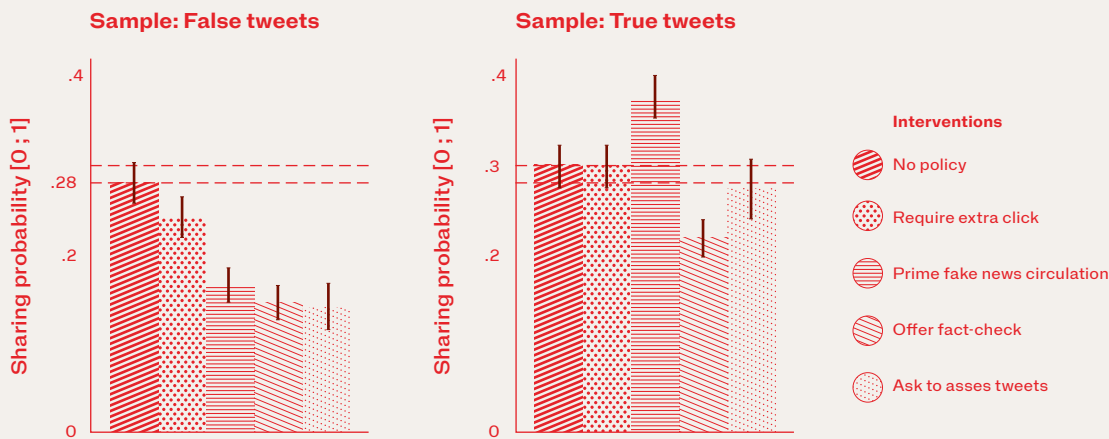
How can we encourage people to think before they share?

How effective could these various interventions be? What mechanisms do they activate? A recent experimental study on 'Curtailing False News, Amplifying Truth' provides some answers. Conducted by Sergei Guriev, Émeric Henry, Theo Marquis and Ekaterina Zhuravskaya during the 2022 mid-term legislative campaign in the United States, it used different treatments to assess their impact on the circulation of both false and true information. The study exposed 3,501 American X/Twitter users to four political news tweets: two containing misinformation and two containing facts.

The participants, who had to decide whether or not to share one or more of these tweets on their X/Twitter account, were randomly divided into groups to receive different treatments. In the first group (the *No policy* control group) they could do whatever they wanted with these four tweets. In the second group (*Require extra click*), they had to click one more time to confirm their sharing decision — a slightly more tedious process. In a third group (*Prime fake news circulation*), they received a 'nudge' message prior to sharing, inspired by the incentives proposed by

EFFECTS OF VARIOUS TREATMENTS TO COMBAT MISINFORMATION

Sample: False tweets

Sample: True tweets

Interventions

- No policy
- Require extra click
- Prime fake news circulation
- Offer fact-check
- Ask to asses tweets

Source: S. Guriev, É.Henry, T. Marquis and E. Zhuravskaya, 'Curtailing False News, Amplifying Truth', *CEPR Discussion Paper*, No. 18650, 2023. https://cepr.org/publications/dp18650

Pennycook and Rand: 'Please think carefully before retweeting. Remember that a significant amount of fake news circulates on social networks.' The fourth group, *Offer fact-check*, were informed that two tweets contained false information detected by PolitiFact.com, a well-known fact-checking non-governmental organisation. They were given the link to access the fact-check. At the end of the survey, all participants were asked to rate the veracity and partisan tendency of each post.

The figure above illustrates the effects of the different treatments on the sharing of false information (left-hand panel) and true information (right-hand panel). It shows that all the treatments helped to reduce the rate of sharing false information. In the *Require extra-click*, *Prime fake news circulation* and *Offer fact-check* groups, the sharing rates were respectively 3.6, 11.5 and 13.6 points lower than in the control group, bearing in mind that 28 per cent of the latter's members shared one of the tweets containing false information. However, not all the interventions had the same effect on the rate of sharing true information, which was 30 per cent in the control group: asking for an extra click before sharing had no discernible effect; offering access to a fact-check reduced the sharing of truthful tweets by 7.8 percentage points; but sending a behavioural warning message (*Prime fake news circulation*) increased the average rate of sharing truthful tweets by 8.1 points.

All these results establish a clear hierarchy of the effectiveness of policies designed to improve the accuracy of shared content. The *Prime fake news circulation* policy, which encourages users to think about the consequences of sharing false information, appears to be more effective, as it encourages the 'sharing discernment' advocated by Pennycook and Rand: it increases the sharing of true information while decreasing the sharing of false information.
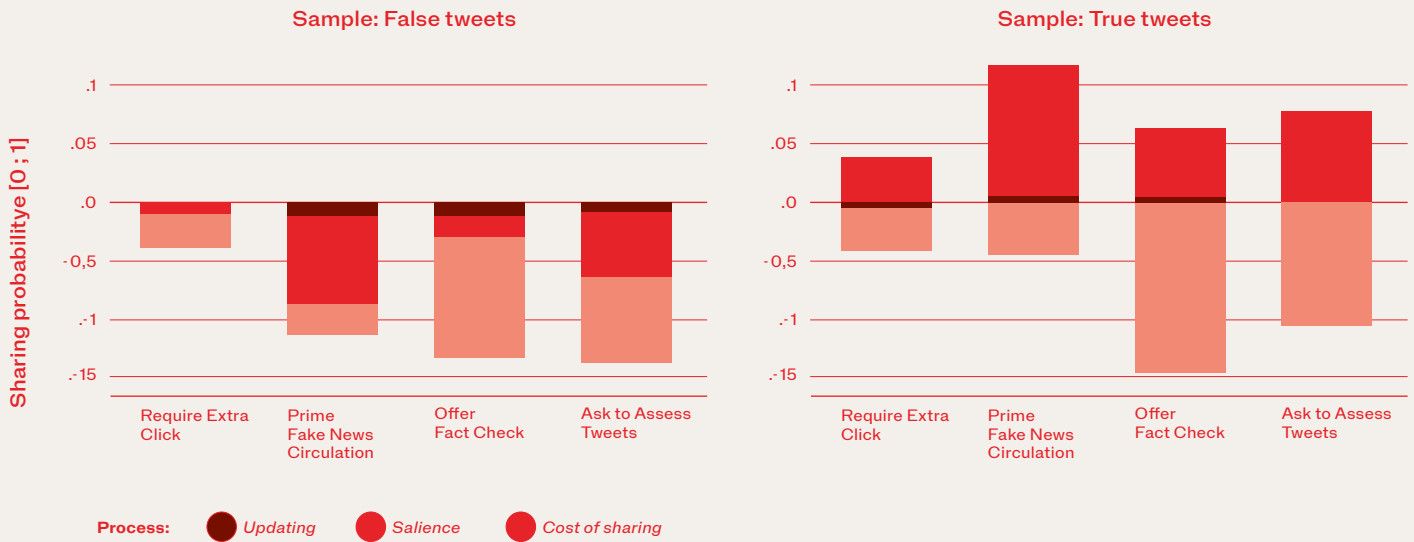
The major impact of reputation effects

To understand the mechanisms underlying the differentiated effects of these treatments on the sharing of true and false information, the study looked at the motives that encourage users to share information on social media. It shows that the perception of veracity reinforces the sense that sharing information is useful for reputational reasons, that's to say not wanting to appear ill-informed in the eyes of one's audience. Information matching the user's opinion also increases feelings of satisfaction when sharing it, be it to convince an audience or to signal political identity.

The study confirms that it is possible to influence sharing through three processes: updating, salience and cost of sharing. The first process leads users to revise their beliefs about the veracity or partisan alignment of content. For example, exposure to fact-checking aims to change one's perception of information accuracy. The second process increases the salience of reputational concerns over partisan motives, so that the user pays more attention than before to the

BREAKDOWN OF THE EFFECTS OF DIFFERENT TREATMENTS TO COMBAT MISINFORMATION

Sample: False tweets

Sample: True tweets

Sharing probability [0 ; 1]

.1 .05 .0 -0,5 -.1 .-15

Require Extra Click · Prime Fake News Circulation · Offer Fact Check · Ask to Assess Tweets

Process: ● *Updating* ● *Salience* ● *Cost of sharing*

Source: S. Guriev, É. Henry, T. Marquis, and E. Zhuravskaya, 'Curtailing False News, Amplifying Truth', *CEPR Discussion Paper*, No. 18650, 2023.
https://cepr.org/publications/dp18650

The perception of veracity reinforces the sense that sharing information is useful.

veracity of information when deciding to share it. Treatments that encourage caution (*Prime fake news circulation*), for example, are designed to affect this salience. The third process, which consists of requesting an additional click for confirmation, regardless of information veracity, increases the cost of sharing for the user. Each process affects this cost.

The figure below breaks down the effects of these three processes. Surprisingly, treatments designed to revise beliefs about the veracity of information, such as fact-checking, have little impact. In fact, the overall effect of each treatment stems from a combination of the salience of reputational concerns, partisan motives and the cost of sharing. Salience in particular explains the difference between the effects of the treatments on the sharing of true and false

information. Improving (or protecting) one's reputation increases the sharing of true information and reduces the sharing of false information. All the treatments, to varying degrees, increase salience, with the message encouraging caution (*Prime fake news circulation*) having the greatest effect. At the same time, the friction associated with the different treatments reduces the sharing of both true and false information. The additional costs of the *Prime fake news circulation* treatment are considerably lower than those of the *Offer fact-check* treatment, which makes this type of intervention more effective in increasing the sharing of true information.

A question of efficiency

The results of this study have two implications for policies aiming to fight misinformation. First, they confirm the effectiveness of short-term actions to encourage users to think about the consequences of circulating false information, as recommended by Pennycook and Rand. This method reduces the sharing of false information and increases the sharing of true facts, without reducing the overall engagement of social media users. Second, these results show that with fact-checking users share less false information, not because they discover that it is false, but because at the moment of sharing they become aware of the need to check the veracity of the information. As a result, despite involving

Sign outside a hot dog restaurant in Chicago, after the televised presidential debate on 11 September 2024, during which Donald Trump claimed that in Springfield, Ohio, immigrants were stealing people's dogs and cats to eat them.

significant investment, fact-checking by professional verifiers could be less effective than fact-checking by an algorithm, which is faster (occurring earlier in the sharing process) and less costly, but more prone to error. In the latter case, the user is quickly informed that the content was flagged as suspect by the algorithm, heightening concern for veracity.

These short-term policies are obviously complementary to, and not a substitute for, long-term policies such as digital literacy. The study also shows an interesting mechanism that underscores this complementarity: if the users, concerned about their reputation, know that their audience is more alert to misinformation as a result of better education, they are less likely to spread misinformation. However, short-term policies are likely to foster habituation, which may reduce their effectiveness. It might be wise to use them only during periods of heightened risk, such as election campaigns.

📖 REFERENCES

→ Barrera, O., Guriev, S., Henry, E. and Zhuravskaya, E., 'Facts, Alternative Facts, and Fact Checking in Times of Post-truth Politics', *Journal of Public Economics*, 182, 2020, pp. 104–123.

→ Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J. and Sircar, N., '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 of the United States of America*, 117 (27), 2020, pp. 15536–15545.

→ Guriev, S., Henry, É., Marquis, T. and Zhuravskaya, E., 'Curtailing False News, Amplifying Truth', *CEPR Discussion Paper*, No. 18650, 2023. https://cepr.org/publications/dp18650

→ Henry, É., Zhuravskaya, E. and Guriev, S., 'Checking and Sharing Alt-Facts', *American Economic Journal: Economic Policy*, 14 (3), 2022, pp. 55–86.

→ Nyhan, B., 'Facts and Myths about Misperceptions', *Journal of Economic Perspectives*, 34 (3), 2020, pp. 220–236.

→ Pennycook, G. and Rand, D., 'Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation', *Nature Communications*, 13 (2333), 2022.

→ Persily, N. and Tucker, Joshua A., *Social Media and Democracy*, Cambridge; Cambridge University Press, 2020.