Democratic Efficacy and the Varieties of Populism in Europe

Report

Building an Immune System Against Fake News

Intervention Toolkit to Spot Disinformation

May 2022

Author(s)
Peter Kreko (Political Capital), Gábor Orosz (University of Artois), Laura Faragó (Pazmany Peter Catholic University), and Benedek Paskuj (British Broadcasting Corporation)

Contact Information
contact@demos.tk.hu
kreko@politicalcapital.hu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 822590. Any dissemination of results here presented reflects only the consortium's view. The Agency is not responsible for any use that may be made of the information it contains.

Table of Contents

Part 1: Intervention Toolkit for Young Hungarians to Spot Disinformation

In the first part of this document you can read about a tool that was designed to help Hungarian adolescents and young adults to spot fake news and distinguish them from real news (pages 1-8). In the second part, you can read a detailed report on the scientific background, methodology, and results of the validation of this intervention material (pages 9-40) in two age groups. At the end of the document two appendices can be found regarding the interpretation of the efficacy of the present intervention and the links for the original Hungarian intervention materials.

The brief description below was written to a broader audience who might be interested in implementing the intervention material. This description can serve as a basic background document that can help to take the first steps if you would like to adapt this intervention to other countries or cultural contexts. If you have such intentions, please feel free to contact the designers of the intervention programme. They are more than happy to help in the adaptation process.

The main goal of the intervention was to motivate young adults to read news critically and be vigilant of fake news. The intervention aimed to tap into adolescents' and young adults' core social motives, which call for vigilance against fake news and it uses the basic principles of wise social-psychological interventions (Walton & Wilson, 2018). This intervention used various psychological mechanisms to make long-term changes. Please find a brief explanation of the main psychological mechanisms below.

- Expert role: In contrast to prior fake news interventions, we supposed that youngsters are experts of the online world in which they can easily experience constant learning and satisfy their need to understand. We aimed not only to highlight their competence but to demonstrate how this can be a source of respect and higher status in their peer groups and also by contrasting them with older adults who have higher position and status in many other situations (see Yeager et al., 2016; 2019). This expert role is very important as this can make people motivated to give advice to those who are less competent and in this way we can provide ground for indirect, self-persuasive strategies (Aronson, 1999)
    - We aimed to build on stable, group-based attribution of being vigilant of fake news. Based on a classic study (Miller, Brickman and Bollen, 1975), the intervention material was explicit about the expertise of young adults regarding digital competences. "Young people like you are vigilant of online disinformation." This motivation aimed to change their long-term behaviour.
    - Playing on intergenerational differences (related to generational norms): In the intervention material, we demonstrated that young adults are supposed to be more competent to spot fake news than elderly generations (it might work as a prescriptive norm). They were put in a role in which they were more competent to be vigilant compared to other age groups (it might work as a descriptive norm to be vigilant, Keizer, Lindenberg, Steg, 2008). We not only used this as an alignment of the prescriptive and descriptive norms, but expressed that young adults were similar to their elderly relatives a few years

ago, but by now they could have achieved a significant head start in digital competencies. However, such vigilance has not been the same in the past (dynamic norms, Sparkman & Walton, 2017).

- Sharing fake news is a sign of incompetency that can lead to reduced status and respect: We aimed to implement a mechanism, that demonstrates that although youngsters are competent in the online world, yet they still have to be vigilant as there might be accidents, and when these accidents happen they can have social consequences in terms of reduced in-group reputation. Avoiding these inconvenient situations in the online space is essential (Yeager, Dahl, & Dweck, 2017). Two submechanisms appeared regarding status quo and respect.
    - Ridiculing: This aspect resonates with prior results on reducing conspiracy beliefs by ridiculing those who have conspiracy theories (Orosz et al., 2016). Avoiding ridicule (that results in shame, reduced status and respect) can motivate youngsters to use their brains when they read news.
    - Shame: This is closely related to ridiculing, and also associated with reduced status and respect.
- Prosocial reasons: Based on Grant and Mayer (2009) and Yeager et al.'s (2014) work on prosocial motives, helping a family member can be a good reason to do tedious work and make the effort that is also required for mastering fake news detection. For this reason, the materials motivated young adults and adolescents to give advice in terms of providing relevant strategies to a less competent person who is close to them: parents and grandparents. Emphasising family ties such as in many of the original Aronson (1999) works can be a very powerful persuasive technique especially in the Eastern and Central European setting in which the value of family and security is very prevalent (see Hankiss, 1991).
- Learning is part of the process: The intervention materials emphasised that everybody can improve their digital literacy skills similarly to driving skills. Even young adults might be reminded that Facebook is not the best news source, and they should be aware of reading news with high reputation and with reliable basis.
    - Not spotting a fake news story is like an accident, it is not a trait. In fact, youngsters already have critical skills as they are digitally the most competent generation. However, sometimes for certain reasons (e.g. when they are tired) if they do not practice it, they can make mistakes. In sum, we imported some elements of learning mindset interventions (Yeager et al., 2016, 2019) in terms of effort necessary to progress, good strategy choice can help spotting fake news and asking advice helps a lot in the improvement.

In sum, the present intervention aimed to ring the bell effectively when youngsters face disinformation online that motivates them to use their intellectual capacities. We aimed to prepare indirect persuasion strategies to achieve long-term effects with using the above-mentioned expert role, status-related motives, prosocial reasons, and it aimed to frame digital literacy a competency that everyone can improve with effort, good strategy choice, and advice.

Instructions for the Digital Advice Intervention

As you can see in the second part of this document in details, with using these psychological mechanisms the intervention can effectively help young people to spot fake news in the long run. As the intervention includes indirect persuasion strategies, it is worthwhile for teachers to frame the materials as a questionnaire that gathers advice for the older and digitally less competent family members. Pupils and students should be invited to collaborate and help a university study (these were Pázmány Péter Catholic University and Eötvös Loránd University in the Hungarian version, but please choose a university in the country of implementation). As a cover story the intervention can be framed as a survey that seeks advice from students regarding digital media to assist the older generation using their own words and incorporating their own thoughts.

One of the somewhat unusual aspects of the present work is that teachers or instructors do not have an active role in the intervention. We intentionally chose to provide this relatively neutral role to the teachers. Teachers and instructors are simply expected to provide the students with the appropriate framing (see above) as a class or as homework assignments and to provide a calm setting for completion. The intervention material – once programmed in an online survey format and made available to the students – can be completed by individual students on laptops, tablets and phones. Overall, the content of the intervention operates independently from teachers and they are not expected to talk about the content, spend extra time on it, etc. for achieving optimal efficiency. Please note that below we provide the content of the intervention so that it can be easily reproduced. However, programming or otherwise publishing it in an online survey format is the task of those interested in applying it.

In light of the above, we recommend the following call for teachers:
"Everyone has relatives and acquaintances who are less competent in online or computer-related activities. Today, as digitally competent young people, we want to ask for your help in advising older people who are less at ease in the online world. In collaboration with Pázmány Péter Catholic University and Eötvös Loránd University (*you can provide the name of local universities that may want to collaborate with you here*), you can take part in a study in which first we summarise some useful tips for older people who need digital help. As a knowledgeable member of the younger generation, we want to ask you to give advice to the generation of your parents and grandparents so that they can be better informed in the online world." After an introduction similar to this you can present the materials (pages 4-8) to the students.

What should we teach our parents and grandparents about reading the news?

We are developing a new program for your generation of parents and grandparents.

We would like to ask you to help with this development! The program presents scientifically-informed strategies on what to pay attention to in order to filter out fake-news.

Example strategies:
- Be sceptical of the titles
- A closer look at the source of the news
- Read other reports

If we can learn from your thoughts how to present those strategies, this program can benefit your parents and grandparents.

Why did we create the programme?

We are researchers at Pázmány Péter Catholic University, and we created this program because many older people want clearer guidance on how to navigate the world of online news.

However, this requires even more real examples from you, from young people who are very familiar with the online world and who know exactly what messages resonate with your parents and grandparents when it comes to online things.

Even though we know effective strategies to filter out fake news based on well-founded facts, we're not really the experts on how we can explain these messages to them as effectively as possible - it's you.

What strategies are good for dealing with news?

This program presents six scientifically sound strategies to help spot fake news:
- Be sceptical about titles
- Question information that is outrageous
- Investigate the source of the news
- Check the evidence
- Find more articles
- Think about whether the story is a joke

Each strategy can be fully mastered and developed for anyone.

The above strategies have been formulated by independent researchers at Princeton University.

Strategies in the hands of young people

These strategies have been used regularly by more and more young people in recent years, but there would be a great need for the older generation to be able to apply them routinely as well.

In the following, you can learn some more details about these and why young people consider it important to use their brains when reading news.

We gave young people these strategies and asked them what they thought of them. In the following, we would like to share with you some quotes that they have written.

Original strategy:
Be sceptical about the titles: If the shocking statements in the title sound incredible, they probably are.

Rob's opinion:
"I find it very embarrassing for someone to share content whose title feels fake. I'm a little more lenient with my mother's age group, but when it comes to my generation and especially when someone is close to me, I'm even ashamed for them. Last time, I wrote to one of my cousins not to read and click on autopilot, but to use his brain and delete the silly news he shared before he was ridiculed for it. He took it down, and I was proud he posted fewer and fewer doubtful things afterwards. You don't need any special skills to do this. Anyone who can read can pay attention to it, and over time, they can more and more easily filter out things whose titles are fishy already. My parents tell me to be sane when I go to a party ... Actually, that's all I can advise them when they are on their phone!!"

Rob, a 17-year-old high school student

Original strategy:
Question the information that is outrageous: If you are reading something that is enraging or intimidating, ask yourself the question: Was this shared to make me feel that way? If the answer is yes, think carefully about whether you share it.

John's opinion:
"I used to be pretty scared about all sorts of things I read online. For this reason, I have figured out that if a news story comes across with some threatening message, I will stop for a moment and ask myself: John, can this really hurt you now or is it an unfounded nonsense that was written just to make you be scared? Asking this question and thinking about the answer will protect me from being afraid of a lot of unnecessary things. If I even feel a little that a terrifying text was written by someone to make people like me afraid and not because there is an actual threat, it makes me very angry. If they want to manipulate me through fears and threats, it makes me super mad at them because of my history studies."

John, a 21-year-old student (with a major of history and Hungarian)

Original strategy:
Examine the news source: Make sure the story was published by a source whose accuracy can be trusted. If the story comes from an unknown organisation, try to find out more about who they are.

Anna's opinion:
“If I had been asked where I read news a few years ago, I, like most of my friends, would have said, quite naturally, on Facebook of course. That would not be cool at all today. Of course, there are normal news stories on my Facebook as well, but this nauseating stream of information is also full of junk and fake news. I used to read the news from here without looking at where they came from. Today, I see this as drinking from the sewer. In recent years, like most of my acquaintances, I have searched for news sources that are accepted on a Hungarian and international level and can be trusted for their accuracy. Obviously, we always need our brains, but these sites give me a sure starting point and they already make it easier for me to see where and how the articles of the little shady outlets are distorted.”

Anna, a 25-year-old hairdresser

Original strategy:
Check the evidence. Check the author's sources to make sure they are accurate. Lack of evidence or reliance on unnamed experts may indicate false news.

Tim's opinion:
“I’ve always loved understanding the broader questions in life. However, when I was a teenager, I was even less competent in what to look for when I read about broader theories and I was deceived by a lot of fake news. There are articles that describe world-shattering theories using big and smart-looking words. These stories are really interesting at times, they may even seem clear and logical, but even when they reach the thrill of Marvel movies, they often have no basis. Step by step I learnt that I should stop after a few minutes reading and ask myself: Tim, look at it, what does he say? But why does he say that? What is he basing this on? Who are these experts he mentions? Where did he get your pictures from? Can he be another self-appointed opinion leader? And if the answers aren’t right, I’ll put it in the box of forgetful fictions.”

Tim, a 20-year-old waiter

Original strategy:
Find more articles. If another news source does not report the same story, this may indicate that the news is false. If the story is reported by more reliable sources, it is more likely to be true.

Adam's opinion:
“I used to be enthusiastic about sharing shocking news so that all my acquaintances would know first from me what was going on in the world. Then one day one of my older water polo training buddies, who I looked up to very much, told me he saw my post and I should delete it immediately. He said that if you share new and interesting news, they will think you are really informed. If, on the other hand, you share fake news without reason, they will look at you like a dumb kid, even though I don’t think you are. I went back to the pool and was constantly thinking about when I could delete my post. I’ve been sharing a lot of things since then, but before that, I always check several sites to see if it’s really true.”

Adam, a 24-year-old economist

Original strategy:
Think of the story as a joke. Sometimes fake news is hard to distinguish from humour or satire. Make sure the source is not known for parody and that the details and tone of the story suggest that it was just for fun.

Tom's opinion:
“A few years ago, I was deceived by a news item from The Onion and commented on it as if it were true. A few minutes later, one of my friends wrote to me kindly and asked if everything was okay with me... It might have been pretty funny to my acquaintances from the outside, but it was less funny from within. Everyone reads news tiredly and superficially, sometimes everyone believes in nonsense, so accidents can happen. It’s like driving a car, no matter how tired you are, you always have to be sensible. You can be very smart but if you don’t use your brain while reading the news, it’s not worth much. On the other hand, if you pay attention, you don’t have to be a genius to pinpoint the fake news. In any case, this case reminds me to check the weird news, if they are from some fake or joke site.”

Tom, 21, is a computer science and engineering student

Help explain how to read news to older generations!

We collect short letters from digitally educated young people to provide the best examples to the older generations.

As you have seen, these strategies are not complicated things. It is still important to draw their attention to them, which can be developed very effectively with some attention and effort.

In a later programme, we will want to share some letters (anonymously) with members of your (grand)parents' generation. **Therefore, we would like to ask you to write a letter along the lines below.**

Your letter:

• Start by naming your family member who would need this information for example: "Dear Mom / Dad / Grandma / Grandpa!"!

• summarise in a few sentences in your own words the strategies you have just read! (You can find their description below the text box.)

• Then think a little about that

  o What arguments can you use to tell them the most effective of these strategies that you really want in the future!

  o What can you advise them to follow these strategies even when you are not there for them.

• Finally, formulate your letter in a way that brings out the most pervasive arguments and bring thoughts and feelings closest to them about how to spot fake news.

There are no right or wrong answers and please don’t worry about spelling. Just focus on conveying your idea, thoughts and feelings!

You can write your letter here:

Part 2: Reducing Disinformation Online Among Hungarian Youngsters: A Warning Sign Prosocial Intervention

Political Capital (Peter Kreko) in collaboration with **Gábor Orosz** (University of Artois), **Laura Faragó** (Pazmany Peter Catholic University), **Benedek Paskuj** (British Broadcasting Corporation)

Highlights

Novelty:

- First intervention in Eastern-Europe to reduce fake-news accuracy and promote fake news discernment using family-based prosocial values.
- It targets youngsters: high school (Study 1) and college (Study 2) students.
- It measures both immediate and long-term (one month) results based on a pre-registration.

Methodology:

- A novel intervention material was developed based on the principles of "wise" social psychological intervention (that aims to change interpretation or lay beliefs of certain situations)
- It was a randomised controlled trial with assessing intent-to-treat effects
- Besides the main effects of the intervention, we were interested in its differentiated effects on vulnerable subgroups: minorities, females, first-generation college students.

Results:

- *High school sample* (N_{t1}=1558, N_{t2}=904): **Immediate effects on reducing fake news accuracy** (d=0.17), but no long-term accuracy or discernment results. **Among minority students we found non-significant, but relatively large effect sizes** in both fake news accuracy ratings (d=0.61) and in fake news discernment (d=0.52) one month after the intervention.
- *College student sample* (N_{t1}=462, N_{t2}=356): **Immediate (d=0.27) and long-term effects (d=0.23) on reducing fake news accuracy as well as increasing short-term (d=0.24) and long-term discernment responses (d=0.29)**. Subgroup analyses showed **even stronger short-term (d=0.30) and long-term (d=0.45) discernment effects among first-generation students**. Similarly to high school students, the intervention appeared to be **especially effective among minority college students as it significantly reduced their long-term accuracy ratings of fake news (d=1.19)**. F**inally, the intervention led to 2.5 times stronger long-term discernment effects among college students who support the current Hungarian populist government (d=0.51) compared to those students who are against it (d=0.19).**

Discussion and conclusion:

- The present study suggests that it is possible to **effectively utilise our wise intervention** and family-related prosocial values among Eastern-European youngsters to be more sensitive to fake news. These results suggest that there are methods that

can lead to long-term results among college students, especially among minority and first-gen students and also those who support the current government.

Executive Summary of Psychological Mechanisms

The present work was a randomised controlled trial in which students were randomly allocated to the active control or intervention group after clicking on a link. In both groups students read other students' testimonials and they wrote a letter to elderly family members. In the intervention group they wrote a letter about strategies on how to distinguish fake news from real news; whereas, in the control group they wrote about how to avoid embarrassing or inappropriate online behaviours that elderly and digitally incompetent people engage in. The main outcome was related to accuracy ratings of fake and real Hungarian news immediately after the letter writing and one month later (follow-up). In this task, students evaluated both the accuracy of the fake and the real news.

As in previous wise social-psychological interventions, it mobilised (1) prosocial values of students (2) as generation-based experts (3) towards family members with (4) stealthy persuasive methods to distinguish fake news from real news.

Prosocial motivations: Former studies demonstrated that doctors are less motivated to wash their hands for their own health than the health of their patients. Students are more motivated to remain persistent during boring and tedious tasks if they have prosocial reasons to learn (for their family or community). Making an effort for not only the self, but the loved ones can be a good reason to read the news carefully. The present intervention used this motivational source as the students explained how to read news to the elderly members of their generation.

Expert role: In contrast with prior interventions in which participants needed help or in which they felt incompetent, in the present case they were the experts in their family who can give good advice to older family members who are digitally less competent. This role has three sides: the first is related to lowering resistance concerning the material, the second is the opportunity to contribute to the family, the third is the responsibility that comes with the competence.

Family values: The family-related prosocial values appear in classic, cognitive dissonance-based social psychological persuasive strategies that were used to reduce smoking behaviour or motivate young adults to use condoms in the USA (older siblings explain the reasons against smoking and those for condom use to their younger peers). The family-related values might be even more relevant in the Hungarian, in-group collectivist cultural context (see in detail above the sociocultural context).

Stealthy persuasive strategies: First, students could read about testimonials of their peers about situations in which they did not follow a strategy that helps spot fake news. These testimonials demonstrated opinions about potential reasons why it might be useful to use their cognitive capacities when they read news. Subsequently, based on these testimonials, they were asked to write a letter to their family members in which they explain how they can spot fake news. With this method the main goal was convincing the self and with elaborating these ideas, the participants could indirectly persuade themselves, which is much more effective than direct persuasion when the experimenter tells them what to do. Explaining the strategies and not

following them especially if they are the responsible ones in the Family can create dissonance, which motivates them to behave in accordance with their advice.

Wise interventions for students – A general methodological introduction

How do we aim to re-edit students' interpretation on spotting fake news that can lead to long-term reduced accuracy ratings of disinformation? With the present intervention we did not aim to modify the personality of the students or the objective characteristics of the media context, but the interface between their self and the news reading habits (see this approach in detail in Walton, 2014; Wilson & Walton, 2018). On the basis of prior studies if this interface is adaptively modified it can lead to recursive processes that can result in adaptive long-lasting habits leading to long-term and increasing positive behavioural changes (Yeager et al., 2014). Prior studies with students showed that such interventions can generate adaptive beliefs of the nature of their intelligence and challenge seeking (Yeager et al., 2016; Yeager et al., 2019; Paunesku et al., 2015); improved health and psychological well-being (Crum et al., 2013; Borman, Rozek, Pyne, & Hanselman, 2019; see Dweck & Yeager, 2019); and reduced suspension rates (Okonofua, Paunesku, & Walton, 2016). The list of long-term positive effects of wise interventions in education is very long (see Garcia & Cohen, 2014; Harackiewicz & Priniski, 2018; Stephens, Markus, & Fryberg, 2012; Walton, 2014; Walton & Wilson, 2018, Yeager & Walton, 2011); however, the general methodologies of these novel, online, brief wise interventions can be summarised in 10 points that are valid for the present intervention that aimed to reduce

1. **Randomised controlled trial**: Social psychological wise interventions are rigorously designed and implemented randomised-field experiments in which people are randomly allocated to a control group and one or more treatment groups. With this method, they are supposed to have the same initial demographic, socio-economic, and academic characteristics in the treatment and the control groups. The control group is mostly an active control group in which students receive similar exercises to the treatment without receiving the psychological mechanisms leading to the expected change. These interventions can make it possible to obtain unbiased estimates of effects of different practices (Walton & Wilson, 2018). In these interventions, the examination of potential baseline measure differences, overall and differential attrition differences can guarantee unbiased randomisation and group allocation (e.g. Orosz, Walton, & Dweck, in prep).
2. **Strong theoretical and empirical basis.** Before testing the psychological mechanisms in the field, prior laboratory or field experimental work provides support for the existence of the relevant psychological mechanisms in a highly controlled laboratory context (Walton, 2014; Yeager & Walton, Walton & Wilson, 2018). In the present case, we aimed to capitalise on prior published prosocial purpose interventions in the US (Yeager et al., 2014; Paunesku et al., 2015; Reeves et al., 2020) and unpublished results finding that among Hungarian young adults family-related values can motivate persistent learning (Orosz et al., in prep). Using these prosocial values we expect that students are more motivated to use their analytic thinking to read news.
3. **Targeting**. Wise interventions target specific groups using specific timing (Yeager et al., 2016). We expected that similarly to the above-mentioned prosocial purpose intervention, we expected that the intervention might be more effective among certain vulnerable subgroups. This is why we aim to examine the moderating effects of gender, first-gen and minority status.
4. **Norms.** In the intervention materials we use descriptive social norms (as it is, see Goldstein, Cialdini, & Griskevicius, 2008) and align them with prescriptive norms (as it should be, Keizer, Lindenberg, Steg, 2008). Similarly to prior studies (e.g. Yeager et al., 2014; Yeager

et al., 2019), first, we will present qualitative survey results conveying information about how students would implement certain strategies to spot fake news in their own life. These testimonials include the acknowledgement of the obstacle, emphasize resources that one can mobilize to cope with the difficulties, and include subtle suggestions regarding various solutions (considering the guidelines of Walton et al., 2017). The testimonials subtly provide norms, but without generating resistance that might derive from the perceived reliability of percentages or statistics. These stories were supposed to help participants to articulate their responses in the subsequent self-persuasive exercise.

5. **Stealthy persuasive techniques**. Instead of direct persuasion, brief social psychological interventions use stealthy, indirect, self-persuasive techniques and they increase commitment through action. For example, one of the most common exercises is saying-is-believing, in which students are asked to write a letter about their goals to a fellow student detailing and explaining the reasons they find these goals important (our letter is based on the intervention of Yeager et al., 2014). In the present case, with this explanation students were actively engaged with the idea of noticing fake news. They selected personalized arguments for being committed to spot fake news, they connected this idea to their own lives, and this technique treated students as someone who helps their loved ones and this prosocial motivation could create an extra layer of motivation to be committed to spotting fake news. Similar indirect technique made long-lasting and deeply persuasive effects in prior experimental (e.g. Dickerson, Thibodeau, Aronson & Miller, 1992; Stone, Aronson, Crain, Winslow, & Fried, 1994) and intervention studies (e.g. Okonofua, Paunesku, & Walton, 2016; Yeager et al., 2014; Yeager et al., 2016; Walton & Cohen, 2011).
6. **Recursive processes**. Once a new prosocial interpretation on the importance of paying attention to the validity of news content occurs, it can become self-sustaining. For example, if students think that it is their responsibility to make their family members fake news, this can create long-lasting changes in their behaviours. Merging personal and prosocial reasons (Yeager et al., 2014) to spot fake news in combination of basic competences (Guess et al., 2020), can lead to long-lasting motivations to filter fake news and distinguish them from real ones. Potentially this can lead to a "snowball" effect that can be measured that becomes even stronger over time. One seminal study showed that such recursive processes can be responsible for the greater career satisfaction, psychological well-being, community involvement and leadership of students who participated in a social belonging intervention, 7-11 years after receiving it (Brady, Cohen, Jarvis, & Walton, 2020). With well targeted messages the negative cycles (e.g., ignorance of news sources) can be transformed to positive cycles (e.g. habits to check the source). Students might even have completely forgotten the actual content of the intervention over the years, but the well targeted messages can make a change in the interpretation that will be present even years later. With the words of Wilson (2011), a good intervention is similar to an in-depth discussion that can rewrite the narratives of a given topic forever.
7. **Opportunity instead of helping**. The framing of the interventions is designed to provide an opportunity instead of helping. Some prior competence building fake news interventions framed the intervention as something that can fill the digital literacy gaps, or they can provide something that is missing. Wise interventions aim to avoid depicting students as someone who "needs" help. This inferior position could undermine the expected effect. Instead, students have an opportunity to express their ideas about the given topic, for example within the framework of a saying-is-believing exercise (Yeager & Walton, 2011). The present intervention, similar to Yeager et al. (2016) and Yeager et al. (2019) will put students in the role of experts who can provide advice to their older and less competent relatives.

8. **Direct online reach of the students**. Dissimilarly to prior in-person attempts (Blackwell et al., 2007; Orosz et al., 2017), these interventions are online and reach students directly (e.g., Paunesku et al., 2015). It not only allows student level randomisation (contrasting to class level randomisation of in-person interventions), but they can allow scaling-up the interventions relatively easily (Kizilcec et al., 2017). This method can lead to more effective self-persuasion and it can also eliminate potential resistance against the person who conveys the intervention materials (e.g. the instructor that certain students do not like; see Dweck, 2015). Similarly to other online research in which students could provide potentially sensitive data, we fully respected the regulations of the GDPR.
9. **Intention-to-treat analyses.** We analysed the data of all students who were randomised and assess the effectiveness of the intervention treatment condition(s) contrasting to the control group similarly to prior large-scale wise interventions (Yeager et al., 2016; Yeager et al., 2019). The intention-to-treat analyses–which are the golden standard of this sort of interventions–allow to assess the "real-life" effect of the interventions, including results from students who were misallocated (e.g. who were allocated to the control, but also filled out the treatment for some reason), who did not finish the intervention material, or who dropped out between the intervention and the follow-up, etc. This analysis can provide insights about the "real-life" effect of the intervention (for details regarding potential allocation issues, see Puma, 2009).
10. **Preregistration**. Similarly to prior interventions (e.g. Yeager et al., 2016; Yeager et al., 2019), after the exploration studies and the construction of the intervention and control materials, we pre-registered the study on the candidate's and the preregistration included all of the necessary information.

Prior interventions and their effects on fake news detection

Before we started the design of the interventions, we reviewed the existing literature to be aware of the state of the art and to seek those interventions and elements that might be the most useful to implement wise intervention elements that might change adolescents and young adults' interpretation about their motivations to spot fake news. This mini review below lists those elements we observed on the basis of the current literature (until summer 2021) that might contribute to the design of the present intervention work.

As will be detailed later, there are two main forms of interventions trending nowadays: Nudging and Competence building interventions. From a broader perspective, these two forms of interventions are not surprising as we can see these sorts of interventions in education (Duckworth & Yeager, 2015). The main issue with the Nudging intervention is that they do not necessarily lead to changes in the interpretation of the given field, therefore, they can operate as long as the nudge is present (e.g. as long as one sees a graphic warning). Therefore, long-term effects can hardly be expected. On the other hand, competence building interventions require a significant amount of practice and not necessarily make people motivated to make such efforts (see for example, Yeager & Walton, 2011; Walton & Wilson, 2018). In the present intervention, building on the advances of both nudging and competence building interventions our main goal is changing the interpretation of the participating students on why it is important to pay attention to spot fake news and with using prosocial reasons we aim to make them motivated to generate recursive processes to build their competences.

1. Interventions Using Nudging and Priming

Nudges are behavioural interventions, which modify people's behaviour in a predictable way, through their (online) environment (e.g., graphic warnings) (see Kozyreva et al., 2020; Thaler & Sunstein, 2008). Educational (or System 2) nudges represent a special type of nudges (Kozyreva et al., 2020; Sunstein, 2015), which preserve human autonomy and heighten deliberation, and provide additional information to users. Educative nudges can be used as interventions aimed at combating misinformation. The goal of the following interventions are to get people to slow down and reflect on the accuracy/veracity/credibility of news (Pennycook & Rand, 2021).

1.1. Accuracy-nudging Interventions:

There are multiple forms of nudges one can use to make people slow down and reflect on the news content. These interventions combine techno-cognition (smartly designed online environments aimed at combating misinformation, see Lewandowsky et al., 2017) with educative nudges (Kozyreva et al., 2020). The accuracy-nudging interventions build on the assumption that the sharing of misinformation is the result of inattention, and not the purposeful sharing of fake news (Pennycook & Rand, 2021).

For example, an accuracy reminder before reading a headline (respondents were asked to judge the accuracy of a headline) increased the level of truth discernment in participants' subsequent sharing intentions (2.8 times higher media truth discernment than in the control group, Pennycook et al., 2020). In another study, before participants evaluated the probability of sharing a piece of news, they were asked the following question: *"Please explain how you know that the headline is true or false."* decreased sharing of misinformation (Fazio, 2020).

In Lutzke et al.'s study (2019), participants either read four guidelines for evaluating news online (Guideline condition: *"Do I recognise the news organisation that posted the story?*"; *"Does the information in the post seem believable?"; "Is the post written in a style that I expect from a professional news organisation?"; "Is the post politically motivated?"),* or read and then rated the importance of each guideline (Enhanced guideline condition) participants receiving both types of guidelines trusted, liked, and shared fake news less about climate change on Facebook (without influencing the ratings of real news).

In sum, the advantage of these interventions is that they are not lengthy, but cheap, readily scalable, and preserve users' autonomy (Pennycook & Rand, 2021). Furthermore, these attention-based nudge interventions can be implemented on social media platforms easily (Kozyreva et al., 2020; Pennycook et al., 2021). For example, the accuracy-nudging intervention was employed in a large-scale field experiment on Twitter (Pennycook et al., 2021). On top of these advantages, these interventions can reduce the illusory truth effect (or familiarity effect) when participants hold relevant knowledge (Brashier et al., 2020).

1.2. Metacognitive reflection intervention

Another less direct form of nudging interventions belongs to metacognitive reflection tasks. These prompts can be given to participants before a reading task and they can contemplate instances in which they have encountered and used inaccurate information in their own lives (e.g., *"When was the last time you remember relying on inaccurate information while reading?"*). Then they can generate ideas about how they could be more evaluative when reading to avoid being influenced by inaccuracies in the future (Salovich et al., 2021) effective in reducing judgment errors. This sort of intervention is more strongly related to the wise intervention as it motivates people to interpret a past situation that they can connect to the present one.

1.3. Modifying outcome expectations (based on social cognitive theory)

In the third form of nudge-based interventions respondents read a message emphasising the negative consequences of sharing misinformation (educative nudge). Based on the results of Chen et al. (2015), the message altered participants' outcome expectations and decreased the sharing and liking of misinformation on SNS sites.

1.4. Priming critical thinking

It is also possible to use a cognitive task that requires attention and the use of System 2. For example, in Swami et al. 's study (2014), a verbal fluency task and a cognitive disfluency task were applied in separate studies, both that implicitly elicited analytic thinking and both reduced belief in conspiracy theories.

2. Competence-building interventions

Interventions often aim to build skills and competences (e.g. digital media literacy) instead of changing behaviour (what nudges do). Competency-fostering cognitive interventions are called boosts, which either target cognition (e.g., tips for spotting fake news) and/or the online environment (e.g., a pop-up with information) (Hertwig & Grüne-Yanoff, 2017; Kozyreva et al., 2020). Boosts require human's cooperation and engagement, and instead of merely

preserving human autonomy, they maximize agency and autonomy, which is a big advantage compared to nudges (Hertwig & Grüne-Yanoff, 2017; Kozyreva et al., 2020).

2.1. Digital media literacy interventions:

One type of boosts are digital media literacy interventions. These interventions aim to develop the following competencies: evaluation of the source, evaluation of the evidence, and lateral reading (Kozyreva et al., 2020). Kahne and Bowyer's (2017) pioneering media literacy intervention increased the ability to evaluate news accuracy. Guess et al. (2020) simply provided 12 simple tips about how to spot false news on Facebook and aimed to increase digital media literacy. The intervention increased media truth discernment both in the US and India (except for the rural Indian population) (Guess et al., 2020), both for politically concordant and discordant headlines with long-lasting effects (after multiple weeks). The "Bad News Game" as a typical inoculation technique can also build digital media competencies (see later in detail). In this game players take on the role of a fake news producer and learn six different techniques how to make fake news (impersonating people/experts online, using emotional language (e.g. outrage), group polarisation, floating conspiracy theories, discrediting opponents, and online trolling) (Roozenbeek & van der Linden, 2019). This work uses an inoculation technique (see below in detail) with warnings about fake news and pre-exposure to weakened doses of the techniques used in the production of fake news (Roozenbeek & van der Linden, 2019; Van der Linden et al., 2020). This method appears to be effective in reducing the effects of disinformation. These effects were replicated in the study of Basol et al. (2020) in which the game improves the detection of fake news and increases confidence in judgements. Subsequently, it has been replicated across five different language versions (Roozenbeek et al., 2020).

In a less playful competence development programme in the US and the Netherlands, Hameleers (2020) used a combination of fact-checkers and media literacy intervention (using three important recommendations about how to recognise misinformation) that appeared to be effective in spotting misinformation. Furthermore, Banerjee et al. (2017, Study 2) created guidelines about how to effectively identify fake reviews based on their linguistic cues and this competence building method significantly improved participants' ability to recognise fictitious reviews. In a similar study of McGrew et al. (2019), students enrolled in a 'critical thinking and writing' course and learnt how to evaluate the credibility of online sources, significantly heightened their online reasoning skills. Finally, in a recent study, a problem-based undergraduate online course, focusing on the framing of fake news, aimed to address fake news illiteracy with the inoculation technique à significantly decreased belief in misinformation (Scheibenzuber et al., 2021). All in all, these interventions were relatively brief with focusing on certain guidelines that can help to spot fake news and distinguish them from real news in a more or less playful way.

2.2. Inoculation interventions:

Another type of competence building interventions is related to the inoculation ("prebunking" – pre-emptive debunking) technique (see Compton, 2013; McGuire & Papageorgis, 1961; Van der Linden et al., 2020), which pre-emptively combats disinformation with warning people about the potential threat of manipulation, and revealing a manipulative technique used by fake news creators (prebunking). Therefore, participants can "immunise" themselves even before encountering real disinformation (Lewandowsky & Van der Linden, 2021). Inoculation can therefore be considered as a special form of competency-building intervention. Inoculation

interventions are carried out prior to exposure to disinformation. Nevertheless, debunking interventions are applied after encountering the disinformation (see next section for more details).

Inoculation interventions have long-lasting effects (Guess et al., 2020; Maertens et al., 2020; Pfau & Bockern, 1994). Inoculation techniques provide broad-spectrum immunity against disinformation (e.g. inoculation applied in a specific domain successfully reduced belief in disinformation in another domain, Cook et al., 2017; Lewandowsky & Van der Linden, 2021), but some studies found that the effect of inoculation is limited only to specific arguments (see e.g. Zerback et al., 2020), yet inoculation can be the "vaccine" against disinformation (Lewandowsky & Van der Linden, 2021). The above mentioned, playful "Bad News Game" and its replications used this technique (Basol et al., 2020; Roozenbeek & van der Linden, 2019; Roozenbeek et al., 2020; Van der Linden et al., 2020). This technique was also present in prior, pioneering anti-conspiracy studies in which inoculation could reduce belief in conspiracy theories (Banas & Miller, 2013; Jolley & Douglas, 2017). It can be effectively combined with other techniques such as demonstrating scientific consensus concerning certain topics (Maertens et al., 2020). Furthermore, inoculation revealing the astroturfing technique on social media sites could prevent belief in disinformation (Zerback et al., 2020). This method was successful in various, more specific fields. Inoculation messages about climate change all reduced the effect of misinformation (Cook et al., 2017; Maertens et al., 2020; Van der Linden et al., 2017). "Fake-expert" approach, emphasising the historical attempts of the tobacco industry to undermine the scientific consensus (Cook et al., 2017).

2.3. Debunking interventions:

In contrast to the inoculation techniques, debunking interventions come after the fake news. For example, in the study of Yousuf et al., 2021, respondents watched a video containing social norms, vaccine information and used the debunking of vaccination myths. According to the results, the video containing the debunking significantly decreased the effect of disinformation compared to the control group which received the video without debunking.

2.4. Contrasting inoculation vs. debunking technique:

A few attempts have been made to compare the two above-described interventions. MacFarlane et al., (2020) compared the effects of a debunking intervention named as the *tentative-refutation intervention* (after disinformation, another expert drew attention to the lack of evidence for the false information) and an inoculation intervention named *enhanced-refutation intervention* (instead of simply noting lack of evidence, this intervention emphasised the deceptive and misleading techniques used in the disinformation article and explained how those techniques were used to deceive readers). Both interventions were effective in combating misinformation, but the inoculation one with enhanced-refutation was more successful. These results are in line with Jolley and Douglas (2017) findings in which they found the superiority of inoculation over debunking. They found that when anti-conspiracy inoculation is presented prior to the disinformation, it is more effective in reducing belief in disinformation compared to the condition when disinformation is presented first. According to a recent review, accuracy-nudging interventions and inoculation techniques are both effective in combating disinformation (see Bryanov & Vziatysheva, 2021).

The Development of the Warning Sign Prosocial Intervention

Introduction

Although previous fake news interventions so far have proved very promising, there are three aspects which did not receive much attention. Firstly, no intervention studies have been conducted in Eastern and Central Europe, but these would be of particular interest as these nations are more exposed to Russian propaganda than Western European countries (Helmus et al., 2018), and also, state-sponsored disinformation is more widespread among weaker institutional systems (see for example, Krekó and Enyedi, 2018). Secondly, youngsters were not specifically in the focus of these programmes (for an exception see Kahne & Bowyer, 2017), and this age group has never been examined by randomised controlled trial interventions regarding fake news accuracy. Finally, there were very few scientific studies (Guess et al., 2020; Maertens et al., 2020, McGrew et al., 2019; Zerback et al., 2020) that aimed to assess long-term results and only Guess et al. (2020) and McGrew et al. (2019) implemented a follow-up after at least one month. Our main goals were to address these missing aspects with our intervention.

Nudging and Competence Building Interventions

Two types of interventions were mainly used in prior research: interventions using nudging and priming and competence-building interventions (for a review see Kozyreva et al., 2020 and Pennycook & Rand, 2021). *Educational (or System 2) nudges* modify people's behaviour in a predictable way while preserving human autonomy and heightening deliberation (Kozyreva et al., 2020; Sunstein, 2015). Accuracy-nudging interventions build on the assumption that the sharing of disinformation is the result of inattention, and not the purposeful sharing of fake news, so the goal of such interventions is to make people slow down and reflect on the accuracy of news (Pennycook & Rand, 2021). For instance, an accuracy-reminder prior to the evaluation of fake news significantly increased the level of truth discernment in participants' subsequent sharing intentions (Pennycook et al., 2020). Interventions using nudging (Bryanov & Vziatysheva, 2021; Chen et al., 2015; Fazio, 2020; Lutzke et al., 2019; Pennycook et al., 2020; 2021; Salovich & Rapp, 2021) are quite popular and effective as they are not lengthy, but cheap, readily scalable and preserve users' autonomy (Pennycook & Rand, 2021), reduce the illusory truth effect (or familiarity effect) when participants held relevant knowledge (Brashier et al., 2020), and social media platforms can easily implement them (Kozyreva et al., 2020; Pennycook et al., 2021).

Competence-building interventions aim to build skills and competences (e.g. digital media literacy) instead of merely changing behaviour. Competency-fostering cognitive interventions are called boosts, which require human's cooperation and engagement, and instead of merely preserving human autonomy, they maximise agency and autonomy, which is a big advantage compared to nudges (Hertwig & Grüne-Yanoff, 2017; Kozyreva et al., 2020). One type of boost is the inoculation ("prebunking" – pre-emptive debunking) technique (see Compton, 2013; McGuire & Papageorgis, 1961; Van der Linden et al., 2020), which pre-emptively combats disinformation with warning people about the potential threat of manipulation, and revealing a manipulative technique used by fake news creators (prebunking), so participants can "immunise" themselves even before encountering real disinformation (Lewandowsky & Van der Linden, 2021). Inoculations (even in the form of using online games such as "Bad News,

that put the player in the role of the disinformer) can be applied successfully against disinformation (Cook et al., 2017; Basol et al., 2020; Bryanov & Vziatysheva, 2021; Maertens et al., 2020; Roozenbeek & van der Linden, 2019; Roozenbeek et al., 2020; Scheibenzuber et al., 2021; Van der Linden et al., 2017; 2020; Zerback et al., 2020) and conspiracy theories (Banas & Miller, 2013; Jolley & Douglas, 2017). Inoculation has long-lasting effects (Guess et al., 2020; Maertens et al., 2020; Pfau & Bockern, 1994) and provide broad-spectrum immunity against disinformation (e.g. inoculation applied in a specific domain successfully reduced belief in disinformation in another domain, Cook et al., 2017; Lewandowsky & Van der Linden, 2021). While inoculation interventions are carried out prior to exposure to fake news, debunking interventions are applied after encountering the disinformation (Yousuf et al., 2021). Although both types of interventions successfully combat disinformation, the inoculation technique proved more effective (MacFarlane et al., 2021; Jolley & Douglas, 2017).

Another type of competence-building boosts are *digital media literacy interventions*, which aim to improve readers' evaluation of the source, evaluation of the evidence, and lateral reading skills (Kozyreva et al., 2020). Digital media literacy interventions successfully increase the ability to evaluate news accuracy and spot disinformation (Basol et al., 2020; Banerjee et al., 2017; Guess et al., 2020; Hameleers, 2020; Kahne & Bowyer, 2017; McGrew et al., 2019; Roozenbeek & van der Linden, 2019; Roozenbeek et al., 2020; Scheibenzuber et al., 2021; Van der Linden et al., 2020), either combined with the inoculation technique (e.g. Bad News Game (Basol et al., 2020; Roozenbeek & van der Linden, 2019; Roozenbeek et al., 2020; Van der Linden et al., 2020), or with fact-checkers (Hameleers, 2020). Digital media literacy interventions can be brief and effective at the same time: for instance, providing participants with 12 simple tips about how to spot false news significantly increased their digital literacy skills as well as their media truth discernment both in the US and in India, both for politically concordant and discordant headlines (Guess et al., 2020). This increase in discernment proved long-lasting in the United States (but not in India).

Current study

In the current research we tested the effectiveness of a "warning sign" wise intervention that aims to ring the bell effectively when youngsters face disinformation online that motivates them to use their intellectual capacities. Our goal is to make students vigilant to recognise fake news items as well as to be able to distinguish them from real news. Though this intervention is based on a list of six recommendations (adapted from the study of Guess et al., 2020), which aim to strengthen digital literacy competencies that can support this process, we did not primarily wish to inoculate youngsters against fake news by building their media skills, or to prime (implicitly or explicitly) them to be more deliberative in their news consumption, but to tap into their core social motives, which call for vigilance against fake news. Besides demonstrating a list of advice and building competencies, multiple wise psychological mechanisms (Walton, 2014; Walton & Wilson, 2018) are applied that focus participants' attention on the importance of keeping these news reading strategies in mind. These mechanisms include generation-related expert role, motivations related to status and respect, ridiculing, learning orientation, family-based prosocial values, and psychological distancing.

The first wise psychological mechanism is to attribute an expert role to youngsters, supposing that they are experts of the online world. We aim not only to highlight their competence but to demonstrate how this can be a source of respect and higher status in their peer groups and also by contrasting them with older adults who have higher position and status in many other

situations (Yeager et al., 2018). Expert role is expressed through the stable, group-based attribution (Miller et al., 1975) of being vigilant of fake information (e.g. with emphasising that the youth are vigilant of online disinformation and are models for the digitally less competent older generation). We also apply generational descriptive, prescriptive, and dynamic norms emphasising that youngsters are more competent to be vigilant than other age groups (descriptive norm), are supposed to be more competent to spot fake information than older generations (prescriptive norm) and are even more attentive to spot fake news than they were a few years ago (dynamic norm). Ridiculing and shame are also applied to make youngsters more vigilant to fake news: although youngsters are competent and vigilant in the online world, they sometimes share fake information due to inattention, which can have social consequences with reduced in-group status and respect, which are important aspects of youngsters' lives (Yeager et al., 2018). Ridiculing (Orosz et al., 2016) is an effective way to reduce false beliefs, so inducing shame though ridiculing can motivate youngsters to use their brains when they read news. The intervention also emphasises that learning is part of the process: though youngsters are competent, they also have to learn how to detect fake news, but they should be aware of reading news with high reputation and with reliable basis so as to avoid the loss of in-group respect. We also build on prosocial motivations: emphasising family ties (Aronson, 1999) with asking youngsters to explain the strategies to a digitally less competent family member (e.g. a parent or a grandparent) who is close to them. Furthermore, with the saying is believing technique (Higgins & Rholes, 1978) youngsters indirectly persuade themselves, which is much more effective than direct persuasion (Aronson, 1999). Explaining the strategies can also induce hypocrisy with highlighting the distance between the advice given by youngsters and their own behaviour, which also motivates them to behave in accordance with their advice (Aronson et al., 1991; Stone et al., 1994). The intervention also encourages psychological distancing (Trope & Liberman, 2010) at some point: when youngsters read ambiguous information online, they should not only stop reading, but should ask questions from a self-distanced perspective.

Effect sizes

We conducted a priori power analysis, based on the effect sizes of Guess et al. (2020), whose intervention was the most similar to ours. With 80% power to detect an effect of d=0.2 (α=0.05, one predictor for main effect) we needed 787 respondents (see the pre-registration document: https://osf.io/8tgk6; https://doi.org/10.17605/OSF.IO/8TGK6). We found it impossible to gather enough students for testing the original d=0.08 effects as it would require 3200 students in the follow-up after the potential attrition; however, we did our best to maximise the number of students for the follow-up. Furthermore, in high schools and the college study, the pandemic limited the number of students we could initially recruit into the study and increased attrition between the two rounds of outcome measurement – this decreased our statistical power.

Sample

Based on the article of Guess et al. (2020), we carried out the power calculations to demonstrate the short-term results (d=0.17); however, we did not expect to reproduce the long-term (d=0.08) results as the result of the expected high attrition rate (see Orosz et al., in prep and also see the pre-registration document). Study 1 tested the ability of our novel intervention to improve fake news accuracy ratings over short and long term in a high school student sample where high drop-out, but lower attention rates were expected. However, we aimed to carry out a complementary study. Study 2 was conducted on a sample of undergraduates in a credit course where students could receive partial credit if they participated in both the intervention and the

follow-up session. In this case we expected more attentive responses and also smaller dropout than among high school students.

Procedure and Intervention Content

Structure of the intervention and control materials

Welcomed and briefed on the study participants first filled out measures of social media use, bullshit receptivity, and demographics and then proceeded to their randomly assigned condition.

For the *treatment group* the exercise was framed as a contribution to an online media literacy programme developed for the parents' and grandparents' generation. Participants read about six scientifically supported strategies (all adapted from Guess et al., 2020), accompanied by peer testimonials, that could help one spot fake news online and were then asked to compose a letter to a close family member that summarised the strategies and to reflect on the best arguments and advice that would persuade their reader to utilize these strategies in life. The strategies included scepticism of headlines; looking beyond fear-mongering; inspecting the source of news; checking the evidence; triangulation; considering if the story is a joke. The testimonial encouraging triangulation was as follows:

> **Find more articles.** If another news source does not report the same story, this may indicate that the news is false. If the story is reported by more reliable sources, it is more likely to be true.
>
> Adam's opinion:
>
> *"I used to be enthusiastic about sharing shocking news so that all my acquaintances would know first from me what was going on in the world. Then one day one of my older water polo training buddies, whom I looked up to very much, told me he saw my post and I should delete it immediately. He said that if you share new and interesting news, they will think you are really informed. If, on the other hand, you share fake news without reason, they will look at you like a dumb kid, even though I don't think you are. I went back to the pool and was constantly thinking about when I could delete my post. I've been sharing a lot of things since then, but before that, I always check several sites to see if it's really true."*
>
> Adam, a 24-year-old economist

For the *control group* the exercise was framed as a contribution to a social media literacy programme developed for the parents' and grandparents' generation. It was related to practices of the parents or grandparents that the younger generation finds especially embarrassing (such as posting their pictures on Facebook without asking their permission or sharing personal information on the Facebook wall, etc). The structure of the control material was very similar; however, the topic of fake news did not appear. Participants read about six practices violating the norms of online behaviour and were then asked to compose a letter to a close family member that summarised the practices and to reflect on the best arguments and advice that would persuade their reader to avoid these practices online. The practices included using Facebook's feed instead of private messaging; virtual bouquets for birthdays and name days; inappropriate

emoji use; ‘funny’ profile pictures; inadequate device handling during video calls; mass invites for online games.

The full intervention lasted approximately 27 minutes and the follow-up took 8 minutes. The study was approved by the local university’s ethics committee.

Study 2 for college students was identical in its design to Study 1 except for the number of fake and real news stories evaluated (8-8 vs 4-4 in Study 1) and the measurement of conspiracy mentality, which was added for this study only. The intervention was delivered midway through the first semester. The core strategies to spot fake news of the intervention were adapted from Guess et al. (2020).

Analytic strategy

Using OSL regression models, we examine the effect of the condition (treatment/control) on fake news accuracy and media truth discernment scores (calculated as belief in true news minus belief in false news, see Pennycook & Rand, 2021). We run this analysis for the immediate and the longer-run results (one-month follow up).

Our primary analysis will be the intent-to-treat analysis including everybody who were obtained until the end of the survey and provided data about their fake news accuracy. We cannot analyse the responses of those students who dropped out before the outcome measures, as we do not have relevant DV variables (the accuracy of fake- and real news). In the follow-up study, we can only use the data of those respondents who were randomly allocated to the treatment and the control groups, and also finished the accuracy ratings in the follow-up.

Study 1: Secondary school students

Participants

Participants of the intervention were students from various high schools and 1558 students reached the randomised intervention or control materials (M_{age}=16.37; SD_{age}=1.09; 52.20% female; 45.1% male, 2.6% other, 91.40% Caucasian). They had the opportunity to participate in the intervention during their classes in school context. The participation in the study was voluntary. The follow-up data gathering was in the middle of the Hungarian fourth wave of the COVID-19 pandemic, this was one of the reasons why many students were not present in the second follow-up study. To merge data across sessions, we asked participants to provide their initials, day of birth (e.g., 17, without month or year), and the first letter of the first name of their mother. We asked these questions at the beginning of the intervention and at the follow-up, so we could match data between the two. During data cleaning, we identified students by their mothers’ initials and day of birth, and then examined whether they were present in the follow-up. Despite these efforts, because they were missing data on identifiers, some students could not be matched between the intervention and the follow-up. All in all, these reasons led to a dropout of 41% of students and this way we could gather intent-to-treat follow up data from 59% of the allocated students (N=904, M_{age}=16.33; SD_{age}=1.06; 54.80% female; 42.7% male, 2.5% other, 94.60% Caucasian).

Measures

Fake news accuracy and discernment: Following Pennycook and Rand (2019) we captured perceived news accuracy by having participants rate real and fake news items on a 5-point scale. In Study 1 there were 4 real and 4 fake news items, none of them with political content. In Study 2 there were 8 real and 8 fake news items, half of which were political. All news items had been pre-tested in a prior replication study (Kreko, Farago, & Orosz, in prep). Fake news discernment scores were calculated by subtracting the mean perceived accuracy of fake items from the mean perceived accuracy of real items.

Results

Immediate effects of the intervention concerning fake news accuracy ratings

The effect of the treatment on the immediate accuracy ratings, β=-0.17, t(1541)=-3.36, p<0.001, d=0.17 were significant (Figure 1). However, this was not true for the long-term fake news accuracy ratings, β=-0.04, t(354)=-1.36, p=0.217, d=0.04 (Figure 2). Neither the short-term, β=0.07, t(1541)=1.41, p=0.160, d=0.07, nor the long-term discernment effects, β=0.01, t(903)=0.217, p=0.836, d=0.01, of the intervention were significant compared to the control group.

Figure 1. Immediate fake news gullibility (accuracy ratings) ratings of high school students by intervention condition

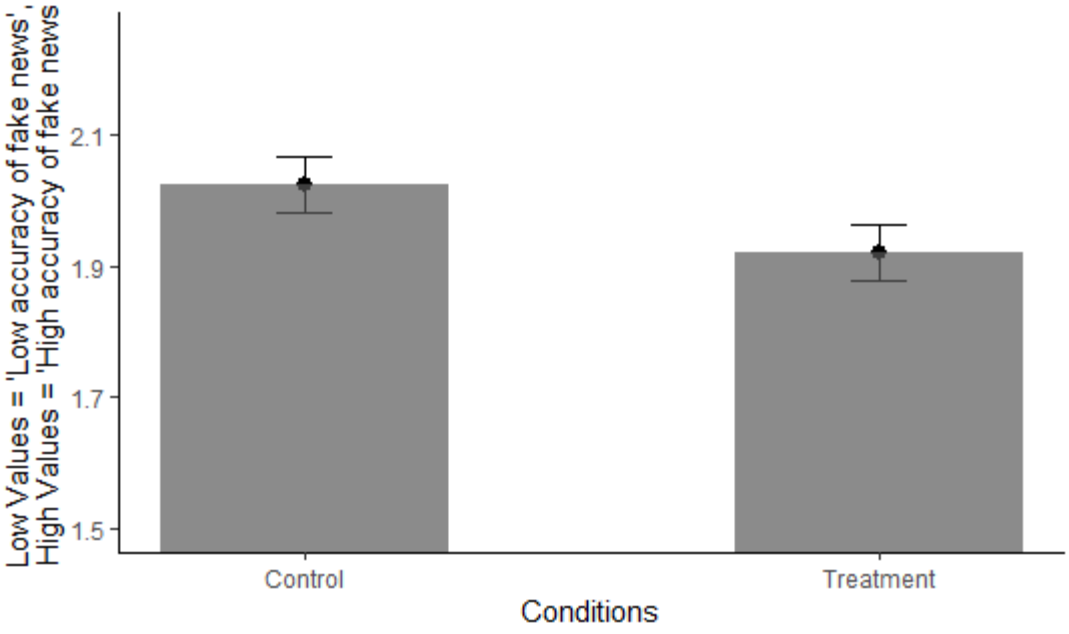
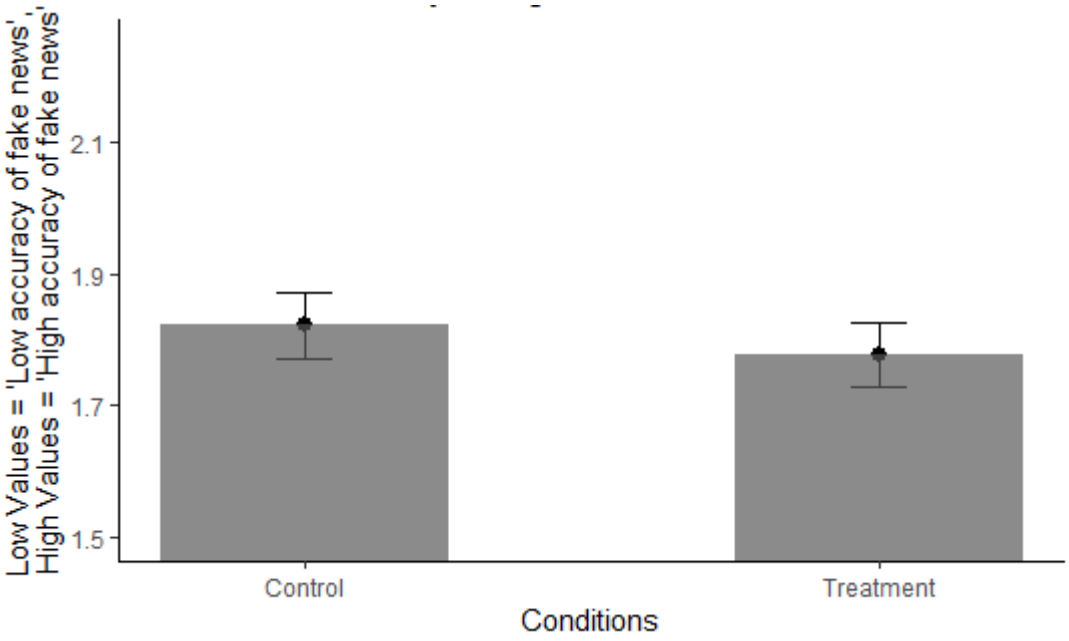


Figure 2. Long-term fake news gullibility (accuracy ratings) ratings of high school students by intervention condition



However, among minority students the intervention appears to be promising as it seems to lead to the expected recursive processes. The effects are larger and close-to-significant in the long-term regarding both fake news accuracy, β=-0.61, t(46)=1.88, p=0.066, d=0.61, and fake news discernment, β=0.52, t(46)=1.601, p=0.116, d=0.52. It appears that this group might require further scientific examination (See Figure 3 and Figure 4).

Figure 3. Long-term fake news gullibility (accuracy ratings) ratings of minority high school students by intervention condition

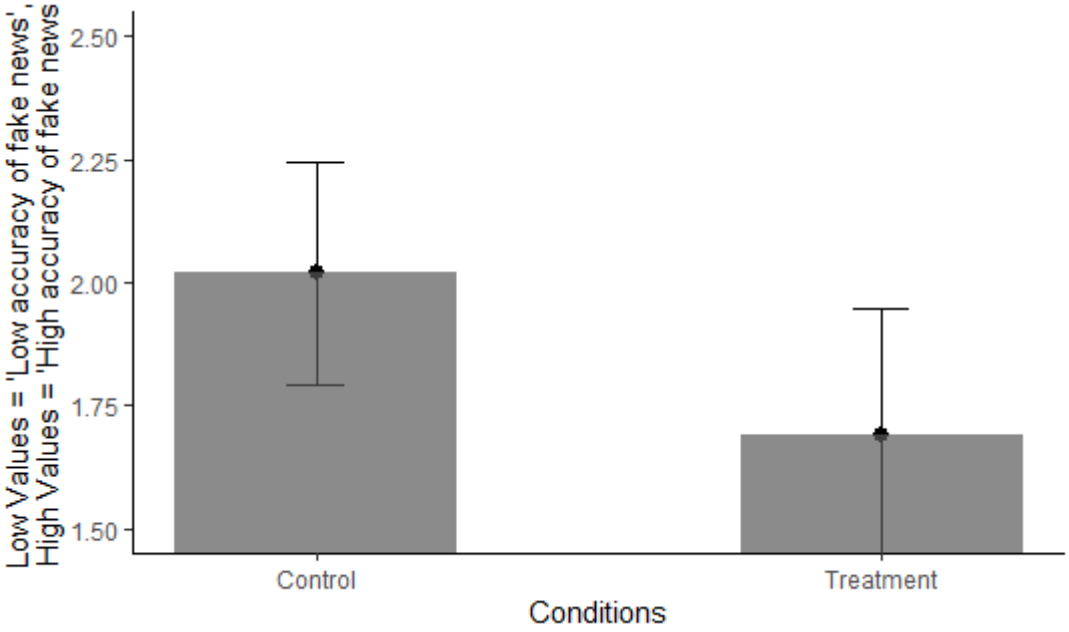
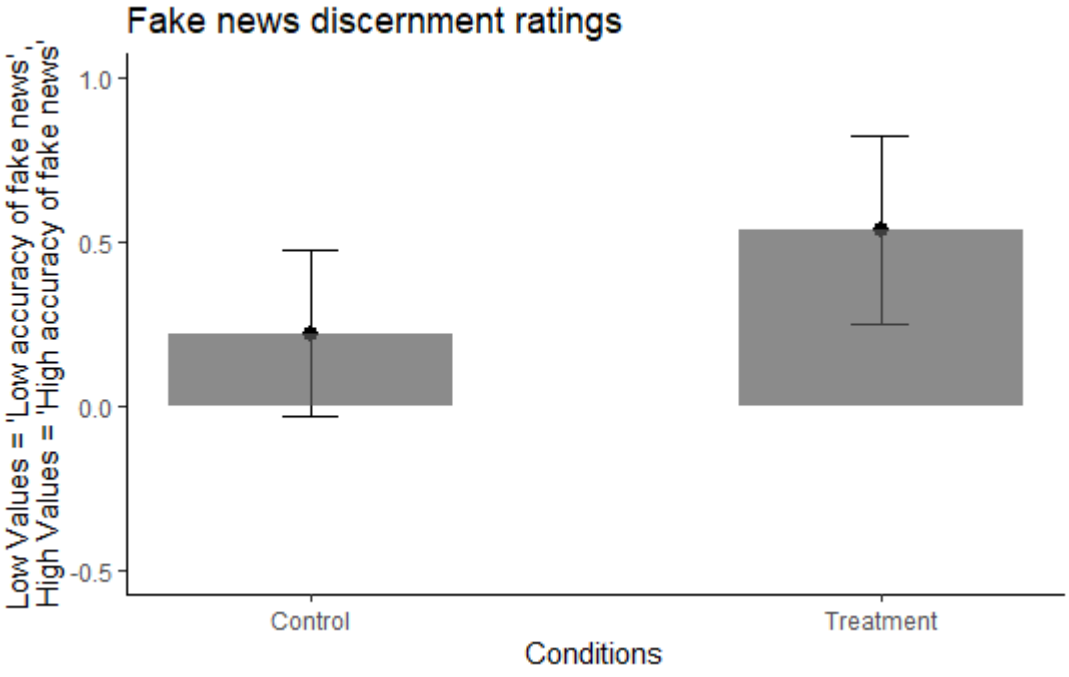


Figure 4. Long-term fake news discernment (the ability of differentiation between fake and real news) scores of minority high school students by intervention condition



Study 2: University sample

Methods

Participants

Participants of the intervention were students with various majors from a Hungarian public university and 462 students reached the randomised intervention or control materials (M_{age}=22.40; SD_{age}=4.82; 73.40% female; 26.4% male, 0.2% other, 94.20% Caucasian). They were enrolled in a mandatory class and, although participation in the study was voluntary, no enrolled students in attendance chose not to participate. The follow-up data gathering was in the middle of the Hungarian fourth wave of the COVID-19 pandemic and it is always possible that students did not provide an appropriate Student ID that could prevent us from joining their follow-up data to their intervention data. All in all, these reasons led to a dropout of 23% of students and this way we could gather intent-to-treat follow up data from 77% of the allocated students (N=356, M_{age}=23.31; SD_{age}=4.46; 75% female; 95.20% Caucasian). There were two students who were randomly allocated to the original intervention or condition but did not finish the follow-up.

Measures

Fake news accuracy and discernment: We used the same, pretested fake news accuracy and discernment measures as in Study 1; however, in this case we added four real and four fake political news to the measure set. Accuracy scores were calculated on the basis of the mean of fake news accuracy scores. Discernment scores were calculated by subtracting the mean of the fake news evaluations from the real news evaluations.

Results

Immediate and long-term effects of the intervention concerning fake news accuracy ratings

The effect of the treatment on both immediate, β=-0.27, t(458)=-3.02, p=0.003, d=0.27, and long-term fake news accuracy ratings, β=-0.23, t(354)=-2.16, p=0.03, d=0.23, were significant. Furthermore, it was also true for the immediate, β=0.24, t(241)=2.58, p=0.010, d=0.24, and long-term, β=0.29, t(354)=2.75, p=0.006, d=0.29, fake news discernment scores, see Figures 2, 3, 4, 5.

Figure 5. Immediate fake news gullibility (accuracy ratings) of university students by intervention condition

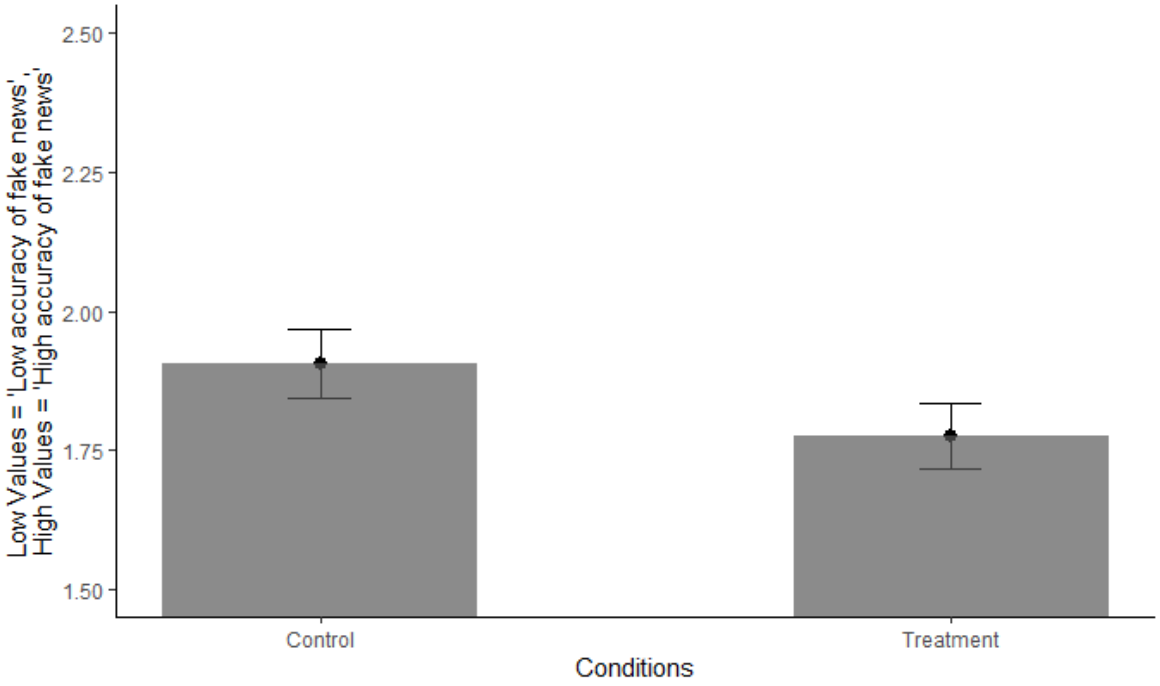


Figure 6. Long-term fake news gullibility (accuracy ratings) of university students by intervention condition

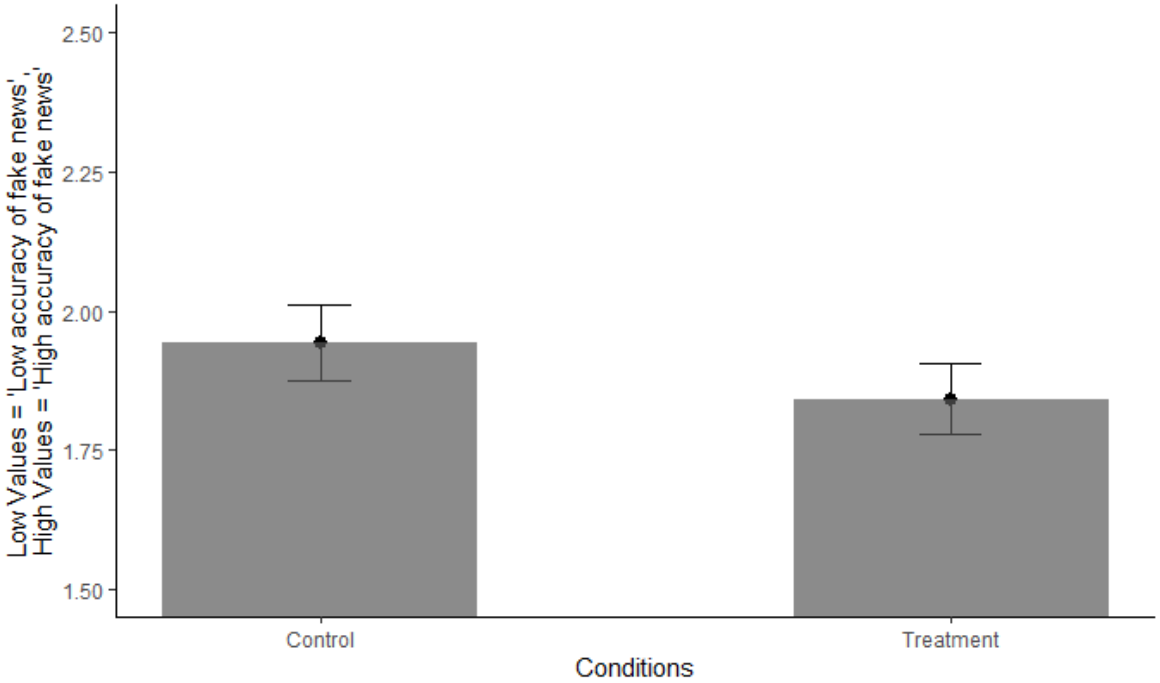
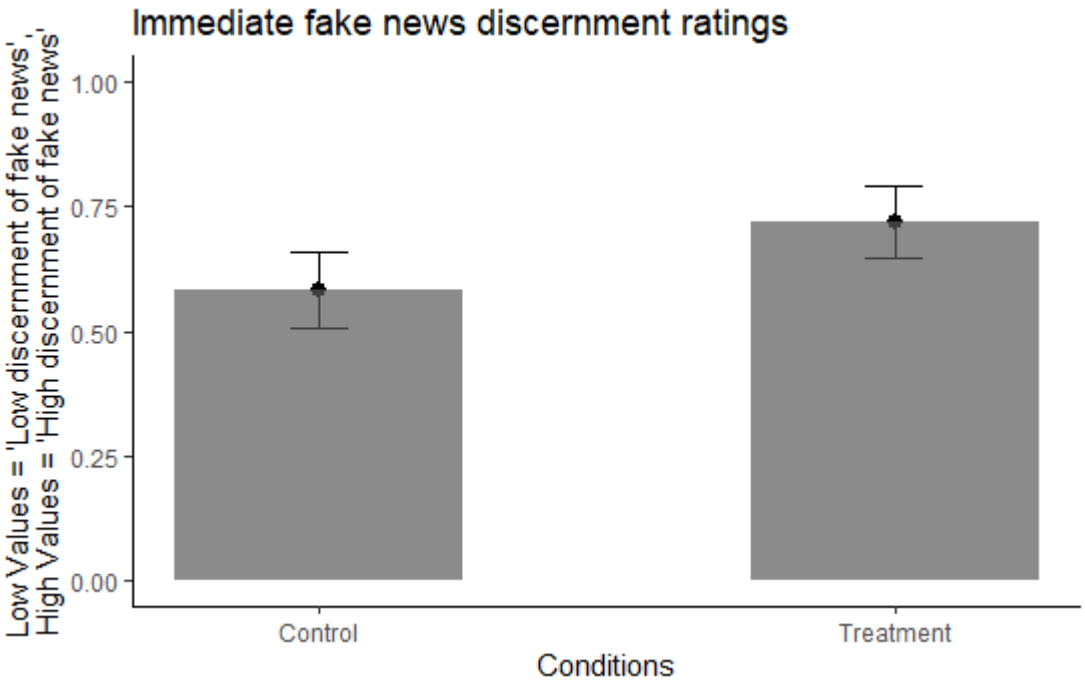
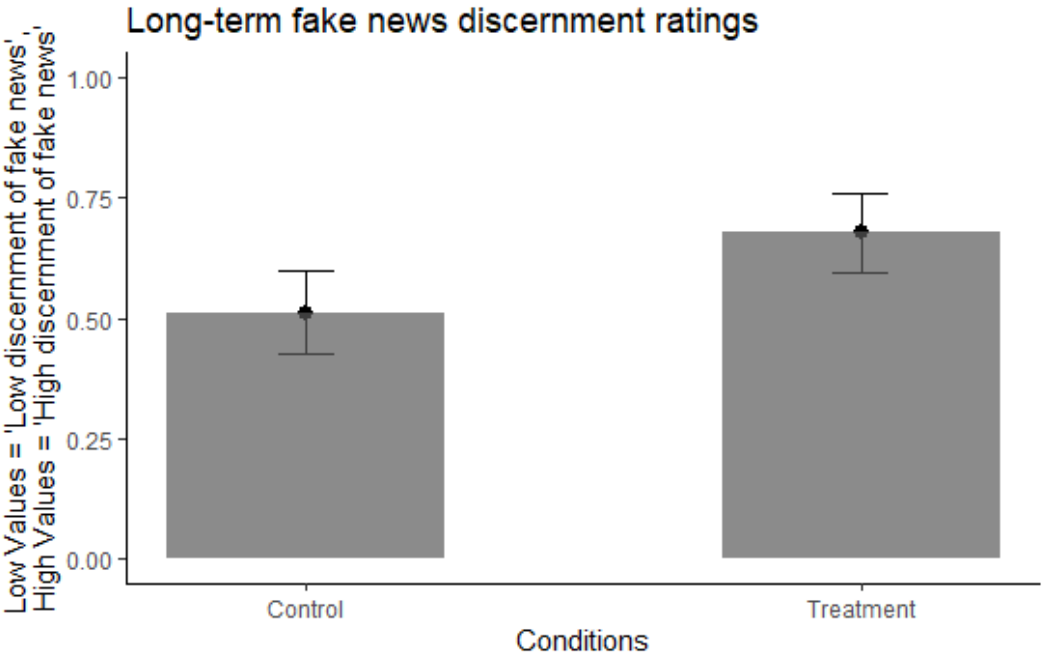


Figure 7. Immediate fake news discernment ratings of university students by intervention condition



Immediate and long-term effects of the intervention concerning fake news discernment ratings

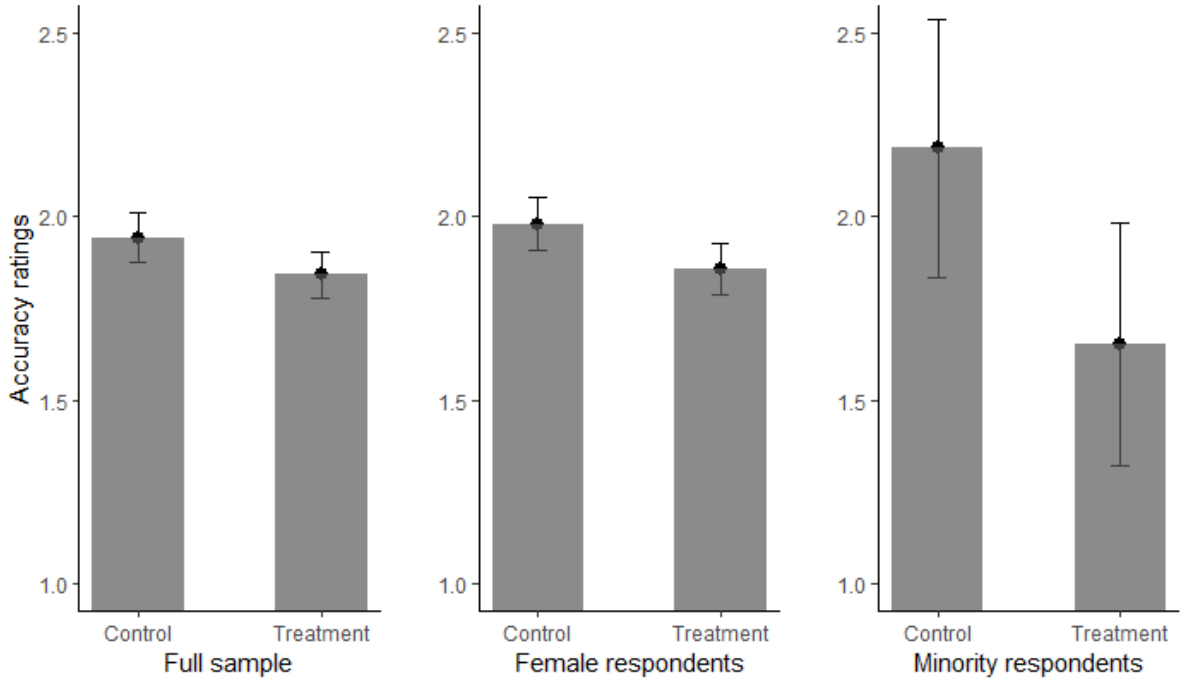
Figure 8. Long-term fake news discernment ratings of university students by intervention condition



We were also interested in whether the intervention had an effect on vulnerable or at-risk groups. In the present report we only focus on female or minority respondents' long-term changes in the *fake news accuracy* ratings. As Figure 5 demonstrates the intervention reduced fake news accuracy ratings among both female, β=-0.27, t(265)=-2.37, p=0.018, d=0.27 and minority β=-1.19, t(15)=-2.17, p=0.047, d=1.19 respondents in a long run.

Long-term effects of the intervention concerning fake news gullibility in subgroups

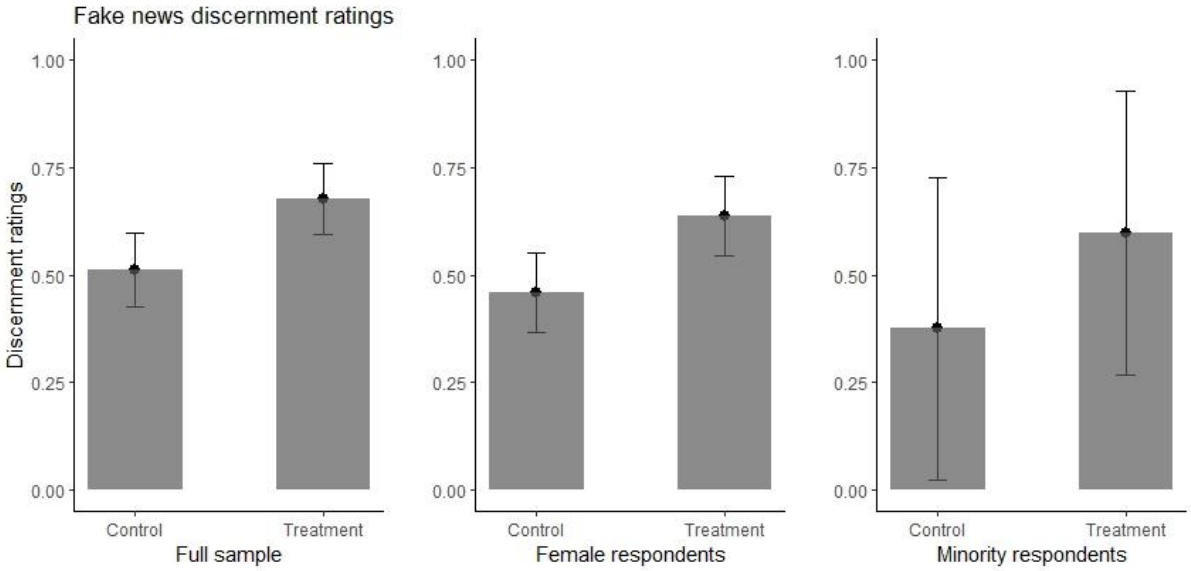
Figure 9. Fake news gullibility (accuracy rating) differences between the intervention and the control groups in the full sample (left panel), among female participants (middle panel) and among minority college students (right panel).



As a next step we were also interested in *fake news discernment ratings* in vulnerable groups such as female and first-generation students. As Figure 6 demonstrates the intervention reduced fake news accuracy ratings among both female, β=0.31, t(265)=2.66, p=0.008, d=0.31 and first-generation β=0.45, t(121)=2.29, p=0.024, d=0.45 respondents in a long run. However, the long-term fake news discernment results did not reach a significant level among minority students because of their low number in the present sample (d=0.39, p=0.38).

Long-term effects of the intervention concerning fake news discernment in subgroups

Figure 10. Fake news discernment differences in the full sample (left panel), among female students (middle panel) and among minority students (right panel).



Fake news discernment and partisanship

We were interested in assessing the effect of the intervention fake news discernment along partisanship. We did not have a specific hypothesis regarding whether the intervention is more effective among pro-governmental or opposing students. Figure 7 and Figure 8 depicts the immediate (p=0.35) and long-term (p=0.20) non-significant interaction results between the condition and the partisanship. However, the subgroup analysis revealed that the intervention appears to increase more efficiently the pro-governmental students fake news discernment, β=0.40, t(110)=2.12, p=0.036, d=0.40, than the students who are opposing the current government, β=0.19, t(346)=1.88, p=0.062, d=0.19, immediately and also in the long-term (pro-government: β=0.51, t(91)=2.49, p=0.014, d=0.51; opposition: β=0.20, t(261)=1.67, p=0.096, d=0.20, see Figures 7 and 8).

Figure 11. Fake news discernment differences along condition and partisanship - short term results

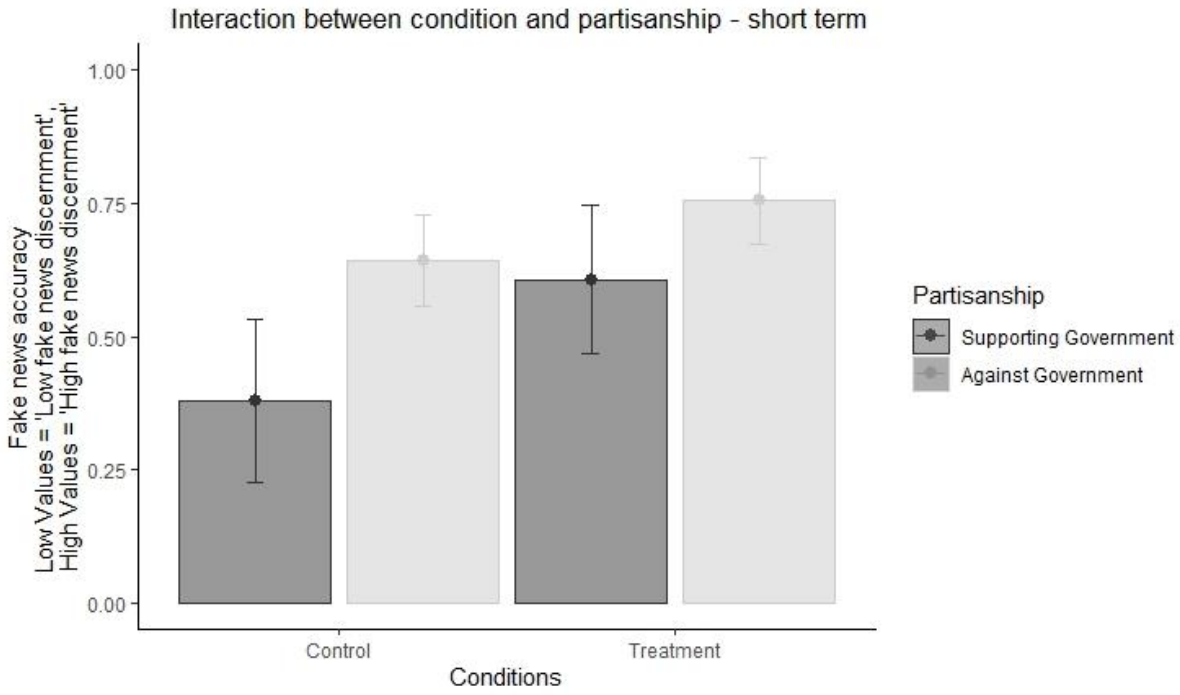
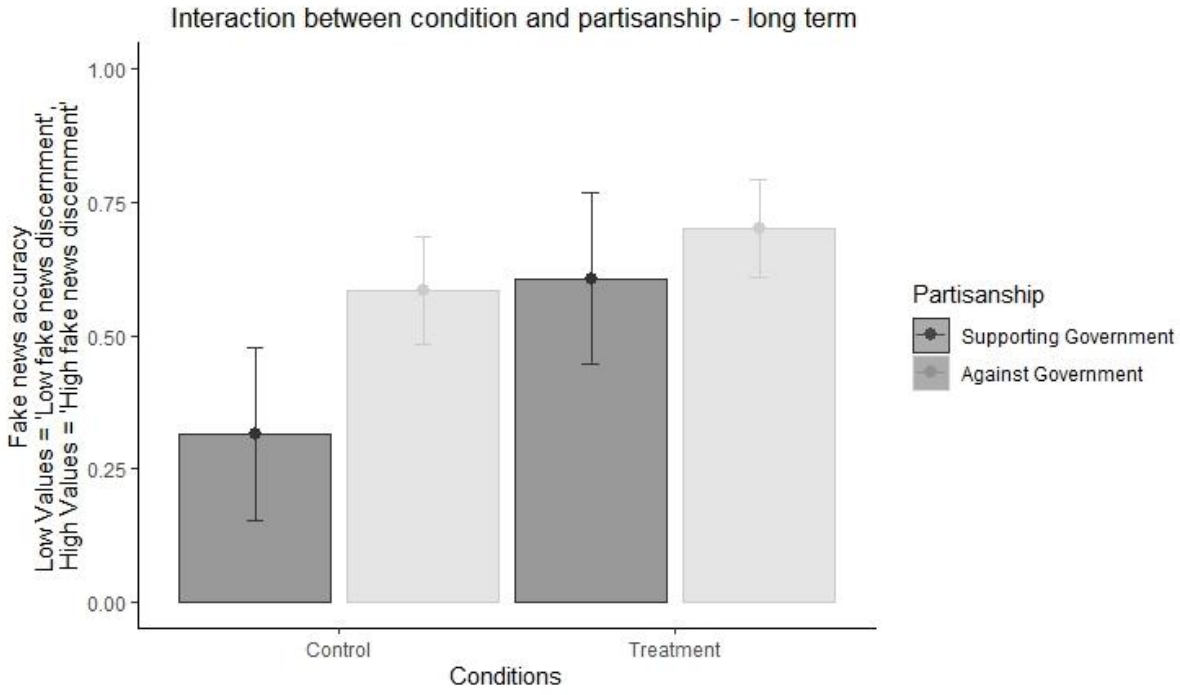


Figure 12. Fake news discernment differences along condition and partisanship



Discussion and conclusion

Anti-scientific and pseudo-scientific views are mainly carried by pseudo-news and conspiracy theories (Krekó, 2018), which are dynamically growing across Europe (McCright et al., 2016). This trend is in line with the declining trust in scientists and scientific institutions (Funk, 2017), but some political trends are also conducive to the proliferation of anti-scientific views: for instance, research shows that countries where populist (left and right) parties have been more successful in elections have higher rates of anti-vaccination rates (Kennedy, 2019). Fake news spreads on social media faster and more effectively than real news, mainly because of its novelty value and its ability to evoke emotions from the recipients (Vosoughi et al., 2018), therefore, vigilance against fake news has never been more important when reading online news.

Our research is conducted in Hungary, an Eastern European country which can be described as having undergone "democratic backsliding" in the past decades (Bozóki & Hegedűs, 2018; Krekó & Enyedi, 2018). The systematic disinformation campaigns initiated by the illiberal Hungarian government (Barlai & Sik, 2017; think tank report of Juhász & Szicherle, 2017), and Russia's soft influence in social media (Helmus et al., 2018) using online astroturfing techniques (Zerback et al., 2020) pose serious danger to the democratic institutions. Therefore, this context can be relevant in terms of strong exposure to governmental propaganda which is not a harsh promoter of analytical thinking either in general or more specifically regarding the evaluation of news contents. Nonetheless, despite its utmost importance, no fake news interventions have been conducted in this context previously.

Furthermore, apart from the contextual shortcomings, there were very few scientific studies (Guess et al., 2020; Maertens et al., 2020, McGrew et al., 2019; Zerback et al., 2020) that aimed to assess long-term results of fake news interventions. Finally, adolescents (high-school- and university students) were not specifically in the focus of these programmes. The main goal of this study is to address these missing aspects with our intervention.

Summary of the main findings

Our intervention was tested among high school and university students with a one-month follow-up via a behavioural fake news recognition and discernment task. In both groups we received the pre-registered and expected short-term effects of recognising fake news. However, among the university students (where the dropout rate was much lower) the follow-up fake news discernment results were not only significant (d=0.29), but even stronger than the short-term ones (d=0.24). It appears that in the group of college students this intervention catalysed processes that made them motivated to spot fake news items and distinguish them from real news items. Even more interestingly, this increased effect derived from vulnerable groups: female, first-generation and minority students. Female respondents' immediate (d=0.29) and long-term fake news discernment (d=0.32) scores were higher in the intervention group compared to the control. Furthermore, among first-generation students, we found an even more salient difference between the control and the intervention groups regarding the immediate (d=0.30) and the long-term (d=0.45) fake news discernment scores. In the case of another, very underpowered subgroup – students who identified as members of a minority group – we found that long-term accuracy ratings of fake news items dropped drastically (d=1.19) compared to the control group.

Theoretical and practical implications

The uniqueness of the present intervention is that it did not aim to nudge people to spot fake news, and it did not explicitly motivate them to build digital competences (see regarding alternative methods Kozyreva et al., 2020 and Pennycook & Rand, 2021). Furthermore, it does not inoculate certain skills for the purpose of being stronger when they read fake news. Prior interventions are mainly individual-focused in which people are motivated to develop their competencies, or develop strategies for themselves, and these prior methods lack the social aspects. The present one capitalised on prior studies (e.g. Grant & Mayer, 2009, Yeager et al., 2014) to promote these *prosocial motivations*. One of the most interesting qualitative aspects of the letters students wrote was related to directly asking their grandparents to call them if they are uncertain regarding any news content. This might not only increase the competency of the young adults through the saying-is-believing exercise (Aronson, 1999) – which is missing content in a prior intervention where the participants are mainly framed as someone who should learn something to avoid being incompetent –, but it also opens the door to use two persons' cognitive capacities to interpret one item of questionable news content. Such interventions can open the door towards intergenerational discussions in which both the younger and the older people can become more competent. Therefore, even longer-run follow-up studies might be relevant and promising.

Another aspect is related to the *cultural context*. It was not arbitrary to choose family-related prosocial behaviours in the Hungarian context. One reason was reaching back to the original condom use studies of Aronson in which the saying-is-believing technique was integrated into a family-related context. The other reason was that in Hungary, family and security related values have been critical since they have been measured. Based on various early sociological analyses (Andorka, 1992; Beluszky, 2000; Csite, 2009), the focus on one's narrow communities in Hungarian society has deep historical roots. In the sociological literature, there is a strong consensus about the dominance of communal prosocial goals involving family and close friends over distal and broader societal goals. The family-level atomisation of Hungarian society started in the 1940s and continued in the 1950s when the ruling socialist party successfully banned more than 90% of clubs, unions and any sort of organisation that was not under direct political control (Hankiss, 1990). In brief, Hungarian society has never recovered from this atomisation and the various political systems instead of making a significant effort to reduce this family-level atomisation, used it for their short-term purposes. The present work demonstrates an example of how it is possible to use these psychological forces to make people more vigilant to spot fake news.

Among high school students, the present intervention reached the fake news accuracy effect size we expected on the basis of Guess et al. (2020) and it was also true for the long-term results. However, in the present case we did not have enough power to demonstrate a significant difference between the two groups. In this age group the intervention did not lead to a significant fake news discernment rate and we can only guess the reasons for that. It is possible that the present intervention was not interesting enough for high school students to take it seriously. It is also possible that the intervention equally reduced real news accuracy ratings and not only the fake news ones. It is possible that playing with the intergenerational persuasive techniques can be a double-edged sword among teenagers (Yeager & Dweck, 2017) because their prosocial intentions towards older members of their family can be contaminated with various sorts of resistance towards them. Deeper examination of the qualitative materials is needed to understand the relevant psychological mechanisms that can be accountable for the smaller effect

sizes and also further studies are required in which we can have stronger control over high school students in school context and in which we can motivate them more effectively to pay attention to the materials.

In a smaller sample, among university students, we found not only long-term reduced fake news accuracy ratings, but we also observed a larger fake news discernment effect. This effect was even stronger than the immediate one. Further studies might investigate how and why the recursive processes developed (Yeager & Walton, 2011; Walton & Wilson, 2018). We can also look for mediators that might be responsible for such changes. For example, we made students more aware of why it might be important to be more analytic when they read the news. It is possible that news reading analytic skills developed over the month between the immediate post-test and the follow up.

In terms of practical implications, the most important aspect is related to minority, first-generation and female students. In line with prior "wise" interventions the present one could help those subgroups who "need" these messages "the most" (e.g. who have lower discernment scores than the average). In future studies it is possible to focus specifically on these groups. It is possible that the family-oriented prosocial content was more important to these sub-groups and the coincidence of being digitally more vulnerable and promoting family-based or communal prosocial values coincided and led to the stronger and persistent effects both in fake news discernment scores.

Finally, it might be interesting to mention that the intervention led to a 2.5 times larger effect among government supporting university students compared to their peers who are against the current populist and illiberal government.

This intervention might provide an example of how it is possible to use family values to motivate people to use their cognitive capacities to spot fake news and to be vigilant for real news in a political context which is much less clear than in the Western-European or American ones. In this region where the government use the public televisions, radios and the local news to regularly disseminate news with ambiguous content this sort of vigilance, especially among their supporters might provide ground for some sort of criticism towards these contents.

Strengths and limitations

Study 1 is one of the first in the fake news intervention literature that collected data from underage students in their everyday school environment – and we faced challenges that ecologically valid contexts pose: substantial attrition, exacerbated by the COVID-19 pandemic, and varying levels of dedication to the tasks among the late-teenage population. Re-doing this study, we would implement measures that maximise attention and serious engagement, especially with intervention materials, and further reduce attrition risks. Future studies that manage to keep attrition at similar levels across secondary and tertiary education contexts will be better able to compare and explain differences in our findings between the two. Both our studies recruited participants from educational institutions and cannot represent the population as a whole: future studies should strive to recruit from more rural and less privileged secondary schools, whereas studies looking at young adults above 18 years need to find those not enrolled in tertiary education – a group that is more prone to believe in fake news (Georgiou, Delfabbro, & Balzan, 2020). The present work did not focus on individual differences in self-reported or behavioural assessments of analytic thinking, and it did not put emphasis on individual

differences in social media use and behaviours related to news consumption. These individual differences might matter regarding the effectiveness of the intervention. Future studies might also analyse the qualitative data of the student letters. It is possible that students who provided a more elaborated response benefitted more in terms of fake news discernment. In sum, the present studies can open the door towards interventions that might effectively increase fake news discernment among Central and Eastern European young adults, especially among female and minority students.

References

Andorka, R. (1992). *Társadalmi változások és társadalmi problémák*, 1940-1990. Statisztikai szemle, 70(4-5), 301-324.

Aronson, E. (1999). The power of self-persuasion. *American Psychologist, 54*(11), 875-884.

Aronson, E., Fried, C., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health, 81*(12), 1636-1638.

Banas, J. A., & Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies. *Human Communication Research, 39*, 184–207.

Banerjee, S., Chua, A. Y. K., & Kim, J.-J. (2017). Don't be deceived: Using linguistic analysis to learn how to discern online review authenticity. *Journal of the Association for Information Science and Technology, 68,* 1525–1538.

Barlai, M., & Sik, E. (2017). A Hungarian trademark (a "Hungarikum"): the moral panic button. In M. Barlai, B. Fähnrich, C. Griessler, & M. Rhomberg (Eds.), *The migrant crisis: European perspectives and national discourses* (pp. 147-169). LIT Verlag.

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition, 3*, 2.

Beluszky, T. (2000). Értékek, értékrendi változások Magyarországon 1945 és 1990 között. *Korall-Társadalomtörténeti folyóirat, 1,* 137-154.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.

Borman, G. D., Rozek, C. S., Pyne, J., & Hanselman, P. (2019). Reappraising academic and social adversity improves middle school students' academic achievement, behavior, and well-being. *Proceedings of the National Academy of Sciences, 116*(33), 16286-16291.

Bozóki, A., & Hegedűs, D. (2018). An externally constrained hybrid regime: Hungary in the European Union. *Democratization, 25,* 1173-1189.

Brady, S. T., Cohen, G. L., Jarvis, S. N., & Walton, G. M. (2020). A brief social-belonging intervention in college improves adult outcomes for black Americans. *Science Advances, 6*(18), eaay3689.

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition, 194*, 104054.

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLOS ONE, 16,* e0253717.

Chen, X., Sin, S. C. J., Theng, Y. L., & Lee, C. S. (2015). Deterring the spread of misinformation on social network sites: A social cognitive theory-guided intervention. *Proceedings of the Association for Information Science and Technology, 52*, 1-4.

Compton, J. (2013). Inoculation theory. In J. P. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 220–236). Sage Publications, Inc.

Cook, J., Lewandowsky, S., & Ecker, U. K. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE, 12*, e0175799. https://doi.org/10.1371/journal.pone.0175799

Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: The role of mindsets in determining the stress response. *Journal of Personality and Social Psychology*, *104*(4), 716-733.

Dickerson, C. A., Thibodeau, R., Aronson, E., & Miller, D. (1992). Using Cognitive Dissonance to Encourage Water Conservation 1. *Journal of Applied Social Psychology*, *22*(11), 841-854.

Dweck, C.S., & Yeager, D.S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481-496.

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.* https://doi.org/10.37016/mr-2020-009

Funk, C. (2017). Mixed messages about public trust in science. *Issues in Science and Technology*, *34*(1), 86-88.

Georgiou, N., Delfabbro, P., & Balzan, R. (2020). COVID-19-related conspiracy beliefs and their relationship with perceived stress and pre-existing conspiracy beliefs. *Personality and individual differences, 166*, 110201.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*(3), 472-482.

Grant, A. M., & Mayer, D. M. (2009). Good soldiers and good actors: prosocial and impression management motives as interactive predictors of affiliative citizenship behaviors. *Journal of applied psychology*, 94(4), 900-912.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (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.

Hameleers, M. (2020). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, 1-17.

Hankiss, E. (1990). *East European Alternatives*. Oxford University Press, USA.

Helmus, T. C., Bodine-Baron, E., Radin, A., Magnuson, M., Mendelsohn, J., Marcellino, W., ... & Winkelman, Z. (2018). *Russian social media influence: Understanding Russian propaganda in Eastern Europe.* Rand Corporation.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12*, 973–986.

Higgins, E. T., & Rholes, W. S. (1978). “Saying is believing”: Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology, 14*(4), 363-378.

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology, 47,* 459–469.

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology, 47,* 459–469.

Juhász, A., & Szicherle, P. (2017). *The political effects of migration-related fake news, disinformation and conspiracy theories in Europe.* Political Capital. http://politicalcapital.hu/pc-admin/source/documents/FES_PC_FakeNewsMigrationStudy_EN_20170524.pdf

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal, 54,* 3–34.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322*(5908), 1681-1685.

Kennedy, J. (2019). Populist politics and vaccine hesitancy in Western Europe: an analysis of national-level data. *European Journal of Public Health, 29*(3), 512-516.

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science, 355*(6322), 251-252.

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest, 21,* 103-156.

Krekó, P. (2018). *Tömegparanoia. Az álhírek és összeesküvés-elméletek szociálpszichológiája.* Atheneum.

Krekó, P., & Enyedi, Z. (2018). Orbán's laboratory of illiberalism. *Journal of Democracy, 29,* 39-51.

Lewandowsky, S., & Van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology,* 1-38.

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change, 58,* 101964.

MacFarlane, D., Tay, L. Q., Hurlstone, M. J., & Ecker, U. K. (2021). Refuting spurious COVID-19 treatment claims reduces demand and misinformation sharing. *Journal of Applied Research in Memory and Cognition, 10*(2), 248-258.

Maertens, R., Anseel, F., & van der Linden, S. (2020). Combatting climate change misinformation: Evidence for longevity of inoculation and consensus messaging effects. *Journal of Environmental Psychology, 70,* 101455.

McCright, A. M., Dunlap, R. E., & Marquart-Pyatt, S. T. (2016). Political ideology and views about climate change in the European Union. *Environmental Politics, 25*(2), 338-358.

McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students’ web savvy: An intervention study. *British Journal of Educational Psychology, 89,* 485–500.

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *The Journal of Abnormal and Social Psychology, 62,* 327–337.

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology, 31*(3), 430–441.

Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences*, *113*(19), 5221-5226.

Orosz, G., Péter-Szarka, S., Bőthe, B., Tóth-Király, I., & Berger, R. (2017). How not to do a mindset intervention: learning from a mindset intervention among students with good grades. *Frontiers in Psychology, 8*, 311.

Orosz, G., Walton, G. M.,& Dweck, C. S. (in prep). Using Mindfulness to Help People Implement a Growth Mindset.

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, *26*(6), 784-793.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in cognitive sciences, 25, 388-402.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature, 592*, 590-595.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31,* 770-780.

Pfau, M., & Bockern, S. V. (1994). The persistence of inoculation in conferring resistance to smoking initiation among adolescents: The second year. *Human Communication Research, 20,* 413-430.

Reeves, S. L., Henderson, M. D., Cohen, G. L., Steingut, R. R., Hirschi, Q., & Yeager, D. S. (2020). Psychological affordances help explain where a self-transcendent purpose intervention improves performance. *Journal of Personality and Social Psychology*. *120*(1), 1-15.

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5,* 1-10.

Roozenbeek, J., van der Linden, S., and Nygren, T. (2020). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review, 1*, 1–15.

Salovich, N. A., & Rapp, D. N. (2021). Misinformed and unaware? Metacognition and the influence of inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47,* 608–624.

Salovich, N. A., & Rapp, D. N. (2021). Misinformed and unaware? Metacognition and the influence of inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47,* 608–624.

Scheibenzuber, C., Hofer, S., & Nistor, N. (2021). Designing for fake news literacy training: A problem-based undergraduate online-course. *Computers in Human Behavior, 121,* 106796.

Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin, 20*(1), 116-128.

Sunstein, C. R. (2015). The ethics of nudging. *Yale Journal on Regulation, 32*, 413–450.

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition, 133*(3), 572-585.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440-463.

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges, 1*, 1600008.

Van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in Psychology, 11*, 2928.

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

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science, 23*(1), 73–82.

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science, 23*(1), 73-82.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review, 125*(5), 617– 655.

Walton, G. M., Murphy, M. C., Logel, C., Yeager, D. S., & The College Transition Collaborative (2017). *The Social-Belonging Intervention: A Guide For Use and Customization.*

Wilson, T. (2011). *Redirect: The surprising new science of psychological change.* Penguin UK.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*(2), 267-301.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*(2), 267-301.

Yeager, D. S., Dahl, R. E., & Dweck, C. S. (2018). Why interventions to influence adolescent behavior often fail but could succeed. *Perspectives on Psychological Science, 13*(1), 101–122.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Paunesku, D. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364-369.

Yeager, D. S., Henderson, M. D., Paunesku, D., Walton, G. M., D'Mello, S., Spitzer, B. J., & Duckworth, A. L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. Journal of Personality and Social Psychology, 107(4), 559-580.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... & Trott, J. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374-391.

Yousuf, H., van der Linden, S., Bredius, L., van Essen, G. T., Sweep, G., Preminger, Z., ... & Hofstra, L. (2021). A media intervention applying debunking versus non-debunking

content to combat vaccine misinformation in elderly in the Netherlands: A digital randomised trial. *EClinicalMedicine, 35,* 100881.

Zerback, T., Töpfl, F., & Knöpfle, M. (2020). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society, 23,* 1080-1098.

Appendix 1: Links of the Hungarian Intervention Materials

You can find the research link here:
High school version: https://pszppke.qualtrics.com/jfe/form/SV_8AGMFPSuUp2n0MK
College student version: https://pszppke.qualtrics.com/jfe/form/SV_8AGMFPSuUp2n0MK

Appendix 2: Mini review regarding effect sizes in fake news interventions

A meta-analysis by Banas and Rains (2010) that considered 40 studies with more than 10,000 participants altogether established an effect size of **inoculation interventions** of about *d*=0.43 immediately after the intervention. The present, immediate results appear to be somewhat smaller than this effect size. Guess et al. (2020) found that on a nationally representative sample in the United States (*d*=0.2, false headlines) and a highly educated online sample in India (*d*=0.13) effect sizes that are comparable to our high school results. The inoculation effects of the Bad News game were observed by comparing preintervention and postintervention credibility ratings of various fake-news items (*d*=0.52) average across all items (Basol et al., 2020; Roozenbeek & van der Linden, 2019). The effects were most pronounced for individuals who had been more susceptible to fake-news headlines in the first place (*d*=0.89). Similar effect sizes (*d*=0.60) were found in a replication of Basol et al. (2020). This intervention resulted in various immediate effect sizes in multiple European countries: Sweden (*d*=0.24); Germany (*d*=0.41); Greece (*d*=0.36); Poland (*d*=0.33; Roozenbeek et al., 2020) leading to an aggregate inoculation effect across countries of *d*=0.37. Maertens et al. (2020) used the earlier-described inoculation technique (*d*=0.95 first measure; *d*=0.28 one week later). Banerjee et al. (2017) carried out a linguistic-cue intervention against misinformation leading to large immediate effects (*d*=2.3 for fake reviews), though they do not report the number of participants in the experimental- and the control groups. In the study of van der Linden et al. (2017), the DV was the respondent's pre and post estimate of the current level of scientific agreement on human-caused climate change. Both the forewarning (*d*=0.33) and full inoculation (*d*=0.75) were effective in conferring resistance against the persuasive attack. In the study of Zerback et al. (2020), people who received inoculation about the astroturfing technique were less influenced by astroturfing comments (*d*=0.19). However, even the immunising effect of the refutational-same treatment was only short-lived and vanished almost completely after a 2-week delay. Hameleers (2020) found that exposure to a media literacy message significantly lowered the perceived accuracy of misinformation (*d*=0.08). MacFarlane et al.'s (2020) enhanced-refutation intervention that is similar to the inoculation technique resulted in a *d*=0.24 effect compared to control. McGrew et al.'s (2019) 150 minutes long 'critical thinking and writing' course (embedded in a critical thinking and writing course) improved university students' ability to make sound judgements of credibility (*d*=0.90) (compared to control). Finally, Jolley and Douglas' (2017) anti-conspiracy intervention presented prior to conspiracy (~ inoculation) led to an immediate effect of *d*=0.62 concerning reduction of anti-vaccine conspiracy beliefs compared to control.

Debunking interventions led to somewhat smaller effect sizes compared to the Inoculation ones. For example, Yousuf et al. (2021) gave a video to the participants containing social norms, vaccine information & debunking of vaccination myths that reduced vaccine-related false beliefs (d =0.49, compared to control). MacFarlane et al.'s (2020) tentative-refutation intervention that can be interpreted under the umbrella of debunking methods led to a d=0.16 effects compared to control. Jolley and Douglas' study (2017) using a method in which a conspiracy was presented prior to anti-conspiracy (~ debunking) intervention led to a d=0.31 effect compared to control.

The preselected default option as a frequently used **nudging technique** had a considerable impact on decisions (a meta-analysis by Jachimowicz et al., 2019, produced a medium-sized immediate effect of d=0.68). A more conscious method, the metacognitive reflection prompts of Salovich and Rapp (2021) in which participants overall made more judgment errors in the no-reflection as compared with the reflection condition led to similar immediate effects in multiple studies (Study 2: d=0.83; Study 3: d=0.52). The accuracy-nudging intervention of Pennycook and Rand (2020) on the sharing of fake news found a relatively small fake news discernment effect (d=0.142). There was a comparable effect of Chen et al. (2015) outcome-expectation educative nudge intervention with an effect size of d=017. Finally, the analytic thinking priming intervention (Swami et al., 2014) led to a medium immediate effect (Study 2: d=0.46; Study 3: d=0.49).

In the light of these effects it is easier to understand the present ones. The overall effect of the intervention among high-school students regarding fake news accuracy scores appears to be mediocre. However, the long-term effect, even if it was not significant, does not appear to be small. With a similarly large sample as in the case of Guess et al. (2020) one can expect a significant intent-to-treat effect (around d=0.08). The short-term effects of the intervention among university students appear to be in line with the inoculation, debunking, and the nudging interventions; however, the long-term effects were stronger than what we found in these interventions. The only small-scale intervention (N=67), relatively long (150 min-long), class-level randomised, offline intervention carried out by the experimenters led to a larger effect size five weeks after the intervention (d=0.90). The present one appears to be a relatively short intervention (~15 without pre and post test), that can allow online person-level randomisation without the biases of the experimenters and which shows larger effects among digitally vulnerable groups such as minorities, females, or first-generation students. Among the political psychological applications, it might be worthy to mention the effect size we measured among supporters of the government.