Applied Econometrics - Final Group Work

Mistakes, Overconfidence and the Effect of Sharing on Detecting Lies

by Marta Serra-Garcia and Uri Gneezy

Group G01

Pablo Alejo HUBER - 20.308.623

Simon Tom CHAPPATTE - 20.333.472

Irena STRBANOVIC - 20.324.901

Ryan Max VINE - 20.325.064

Introduction

"Fake news" are fabricated news that intend to spread misleading information. The dissemination of fake news can be done for two different reasons. Either the person shares the news because it matches a belief he/she already has (e.g. ideological or political reason), in which case the person tends to be less critical of the news and to share it more easily. Or the person has no personal motivation that could skew his perception of the news and shares it simply by believing in it. This paper focuses on the second reason. Its main focus is to determine the factors that lead someone to share a video and to analyze the effect of sharing a video on how quickly a lie spreads.

Summary of the experiments

To do so, the author carried out a lab experiment. He asked several people, whom he identified as the "senders", to have themselves recorded and filmed for two 30-second videos, one real and one fake. In the real video, the "sender" was shown a sheet with a New York Times article that he had to read and summarize for 30 seconds. In the fake video, he was given a blank sheet and had to invent a story. The videos were then randomly shown to another group of people, qualified as the "receivers", who had to guess the truthfulness of the videos. In order to incentivize senders to create convincing fake news, they received 10 dollars if the "receivers" believed their fake news and 0 otherwise.

A total of 3 experiments were made. In the first one, ten senders were recruited, five men and five women. Therefore a total of 20 videos were recorded (two each, one true and one false). The 20 videos were then shown in blocks of five to 572 receivers. After each block of 5 videos, receivers had to say how many videos they thought they had guessed correctly. They won 1 dollar if their answer was correct. They were then asked how correctly they thought they answered compared to the other participants. They once again won a dollar if they had the right answer. They also had the possibility to put a "safe bet" on 3 videos of their choice, allowing them to earn 25 cents more per video if they were true.

Two different treatments were used in the first experiment, the prior treatment and non prior treatment. The prior treatment can be characterized as the treatment group and non prior as the control group. In the prior treatment, the receivers knew how many of the 20 videos were true. In the non prior treatment, receivers didn't know the ratio of true videos. This different treatment allows the authors to analyze if this information would help the receivers to detect lies.

In experiment 2, the number of senders was increased to 42 people, 20 men and 22 women. A total of 84 videos were then shot. This larger sample allowed researchers to use a facial recognition system. This system was used to analyze the physical features and expressions of senders when they were lying or telling the truth. The author then used this data to see if the "receivers" were able to perceive certain physical or speaking characteristics that could help them perceive a lie. The instructions were the same as for experiment 1. Eight videos were shown to 1056 receivers who had to say if they were true or false. One of these 8 videos was then taken randomly and if it had been correctly guessed by the receivers he won 1 euro. As with the first experiment, the receivers had to say how many videos they thought they had guessed correctly. If their answers were right, they won more money.

This experiment 2 only had one treatment, the "No-prior" one. The objective of this experiment was to detect if there were any physical features that could help someone to detect lies and to see if those features were interpreted correctly by people.

The 3rd experiment was conducted by the authors to first understand if lies tend to be shared more easily, and second to analyze the effect of sharing on others' beliefs. To do this, they introduced a second group of receivers. The videos used for this experiment were the same as those used for experiment 1.

Four different treatments were used in this last experiment. The first two: the "shared-true" and the "shared-believed", had the same structure as the treatment used in experiment 2 except that an additional step was added. First, Receiver 1 had to watch two sets of 8 videos for a total of 16 videos. In each set, the number of true and false videos was identical. To prevent a person from being more believed due to certain characteristics (such as beauty or age), they also contained the same number of male and female senders. As for the second experiment, receivers 1 had to watch and assess the truthfulness of each video. They then had to pick one of the 8 videos they wanted to share with the Receivers 2. Receivers 2 then had to choose four of the eight videos to watch, knowing which one of the 8 videos had been shared by receiver 1. What differentiated the first two treatments was the motivation of receiver 1 to share the video. In the “Shared true” treatment, receivers 1 was paid if receivers 2 decided to choose his video to watch and that the video was true. In the “Shared-believed” treatment, receivers 1 was paid a bonus if the receivers chose his video to watch and believed the video was true. In both of these treatments Receivers 2 knew the incentives structure of Receiver 1 .

Two other treatments were used during this experiment: the "shared-no-incentives" treatment and the "no sharing information treatment". In the first one, the author removed the link between Receiver 1's payment and Receiver 2's choice of which video to watch. Receiver 1 just had to share a video he found interesting and receiver 2 knew that his decision had no impact on Receiver 1. This treatment allowed to compare whether the impact of sharing a video was the same when Receivers 1 had no incentives to share the video.

In the "no sharing information treatment”, receiver 2 did not know which video had been shared by receiver 1 . He had therefore no incentives to choose one video over others. In this treatment, receiver 2 just had to choose 4 videos to watch and assess among the 8.

Sample:

For this paper, the experimenters had to make two samples for each experiment: one for the senders and one for the receivers. In the table below, you can see the number of senders and receivers for each experiment.

Experiment	Number of senders	Number of receivers
1 (No-Prior)	10	380
1 (Prior)	10	192
2 (No-Prior)	42	1056
3 (Shared-true)	10	384 (R1) and 384 (R2)
3 (Shared-believed)	10	371 (R1) and 371 (R2)
3 (Shared-no-incentives)	10	198
3 (No-sharing information)	10	185

For the experiment 1 and 3, the senders were research assistants or graduate students at UC San Diego (5 men and 5 women). For experiment 2, they were all undergraduate students at UC San Diego (20 men and 22 women). As for the receivers, it’s a random sample of people all coming from Amazon Mechanical Turk (AMT) with an US IP address. Across the three experiments, their age range was from 18 to 84 and they were 39 years old on average.

To avoid bias and endogeneity, the experimenters forced the receivers to participate in a sound test to make sure that they could hear correctly, because if they couldn’t, they would focus on other variables such as the physical traits and gestures of the senders but not listen to the content of the news article itself.

Hypotheses

Hypothesis 1:

The first hypothesis was the following: “*Receivers are not better than chance at classifying videos containing lies and true statements*.” In other words, this hypothesis argues that even if a person wants to share a true video or assess the veracity of a video, she is simply not able to do so better than by chance. This first hypothesis allows to answer the first research question of this paper, namely: "How good are people at detecting lies".If this hypothesis turns out to be true, it could be one of the reasons that could explain why fake news spreads so easily, simply because people are naturally not able to distinguish the truthfulness of a video.

Hypothesis 2:

The second hypothesis separates into three distinct assumptions. The first one states that: "*Receivers are overconfident in their absolute ability to detect lies*." In other words, it hypothesizes that humans tend to be very confident in their personal ability to detect a true video from a false video.

The second assumption is : "*Receivers are overconfident in their relative ability to detect lies*". This assumption hypothesizes that people tend to believe that they are better than others at detecting lies than what they really are.

The last assumption of hypothesis 2 is: "*Men are more overconfident than women* ». This last hypothesis is interesting because, if it turns out that self-confidence is linked to sharing (being overconfident makes you share more) , it could indicate that men tend to share more and therefore spread lies more.

If hypothesis 2 turns out to be true, it could be very useful to understand why lies are shared so easily, especially if hypothesis 1 also turns out to be true. By combining the two, it would mean that people tend to believe that they are good at judging the truthfulness of a video when they are, in fact, not significantly better than chance. There would therefore be at least 50% probability that a person who wants to share real news ends up sharing a fake one without realizing it. This could explain why lies spread so easily. This hypothesis allows us to answer the second question of the paper, which is: "how confident are people about their ability to detect lies"? The answer to this question will allow the author to understand why lies are so shared.

Hypothesis 3:

The third hypothesis is the following: "*Receivers are more likely to share videos that contain a lie than ones that contain a true statement, both when they are incentivized to share a truthful video and when they are incentivized to share a video that they think another receiver will believe."*

It is based on the hypothesis that when we make a lie, we make it more interesting than the truth. Subsequently lies are more interesting than real events. If we are more attracted to fake news because they are more interesting than truthful videos it could explain in part why we tend to share more fake news. Additionally, if we are overconfident in our ability to detect lies, it will add up and increase the probability to share fake news when we want to share a truthful video. This third hypothesis tests the following research question: "do people share fake news more frequently "? If this hypothesis proves to be true, it could help to understand why lies are more shared than truths.

Hypothesis 4:

The fourth hypothesis is the following: " *Receivers are more likely to watch a video and believe it when they learn that another receiver chose to share it with them*."

In other words, if a receiver suggests to another receiver a video that he found interesting the receiver will be more inclined to watch it and to believe it because it was shared with him. This fourth hypothesis tests the following research question: "are videos that are shared more likely to be believed?" If this statement is true, it will mean that we are more likely to share videos that are false (in case hypothesis 3 is true) and to believe them when they are shared to us. Therefore, it means that we are more likely to watch and believe fake news.

Results Presented by Hypothesis

Since the authors based their study on four hypotheses, we will present their findings for each hypothesis so that the results are easily understandable.

Hypothesis 1:

This hypothesis was tested in experiment 1 and 2. In experiment 1, the authors measure whether receivers are good at detecting lies. Experiment 1 gives data on the frequency at which receivers correctly guess the truthfulness of a video (in percentage points). The experimenters were then able to compare these percentages with the frequency at which receivers correctly guess the truthfulness of a video when they randomly assign true or false to the video (= 50%). So if the results about ability found in experiment 1 are significantly higher than 50%, it would mean that hypothesis 1 is false. To test this hypothesis, the experimenters also decided to use monetary incentives to motivate receivers to do their best.

As mentioned in the **Summary of the experiments** part, they distinguish two groups: the No-Prior group and the Prior group. These two treatments were used to analyze whether knowing the rate of true videos before watching them had an influence on Receivers' answer.

The results from experiment 1 are the following: receivers can only detect lies 50.4% and 51.2% of the time in the No-Prior and Prior treatments respectively. We observe that overall, participants are not significantly better than chance (50%) at detecting lies, which confirms the first hypothesis.

In experiment 2, the authors continue to measure whether Receivers are better than chance at detecting lies in the same way as in experiment 1. As a reminder, the only group watching videos in this experiment is the No-prior treatment group.

In this experiment, receivers were able to detect lies 53.2% of the time. This percentage is significantly higher than the ones obtained in experiment 1 and also significantly higher than chance, but only by 3.2%. This can be explained by the fact that the sample size in Experiment 2 (1056 participants) is larger than in Experiment 1 (380 and 192 participants for the non-prior and prior experiments respectively). A larger sample size allows for more accurate and reliable results. Indeed, it reduces the margin of error as well as the standard deviation. The difference between the results of the two experiments is quite small and could therefore be due to the differences in sample size.

The results presented above show that Receivers have basically no ability to detect lies, and confirm the statement made in Hypothesis 1.

Hypothesis 2:

This hypothesis was tested during the two first experiments. Indeed, the authors set up several questions which allowed them to test the absolute and relative confidence of each participant. After watching the videos, the receivers had to assess how many videos they guessed correctly. If their answers were correct, they earned some money. Question a) of the second hypothesis was used to analyze the absolute overconfidence of each participant. These same participants were then asked to guess how accurate their answers were compared to other participants (by placing themselves in a quartile). Here again, the participants earned money if their answers were correct. This time, the authors analyzed the relative overconfidence of each participant. The money was used to incentivize receivers to respond as best as they could. The authors were then able to compare these results by looking at differences between genders, and conclude whether men were significantly more confident than women.

In the first experiment, in the No-Prior treatment, the actual rate of correct assessments is 51% and 50% for men and women respectively. Men and women believed they correctly assessed 67% and 63% of the videos respectively (absolute overconfidence). In the Prior treatment, the actual rate of correct assessments is 52% and 51% for men and women, respectively, and they believe they assessed videos correctly at 66% and 62% respectively (absolute overconfidence). We can see that the results for the Prior treatment do not differ from the ones for the No-Prior treatment. Women and men have the same ability to detect the truthfulness of videos, are overconfident in their ability to do so and men are significantly more confident than women. The results of the first experiment are consistent with the statements of the first hypothesis.

Column (1) and (2) of the table beside include video fixed effects. Columns (3) and (4) include experiment/treatment fixed effects. The variable "Female receiver" is a dummy variable that takes the value 1 if the receiver is a woman, and 0 if the receiver is a man. We can see that the coefficients obtained for females are statistically significant. The coefficient in column (2) represents the difference of overconfidence between men and women. In fact, we see that absolute overconfidence is negatively correlated with the fact that the receiver is a woman, which means that women are less overconfident than men (but they still are overall). The coefficient in column (4) represents the receivers' beliefs about their position in a quartile. The higher the quartile, the lower the overconfidence level. Since the coefficient for female receivers is significant and positive (=0.331), it means that female receivers are more likely to place themselves in higher quartiles, which advocate for a lower level of overconfidence. So these results confirm hypothesis 2 c).

TABLE 1: Some determinants of confidence

Variables	**(1) Absolute confidence (Beliefs about share of correct ratings)**	**(2)**	**(3) Relative overconfidence (Beliefs about quartiles)**	**(4)**
Constant	0.654*** (0.011)	0.695*** (0.012)	2.201*** (0.135)	2.010*** (0.134)
Share of correct ratings	0.016 (0.013)	0.015 (0.013)	-0.367 (0.250)	-0.281 (0.244)
Female receiver		-0.043*** (0.009)		0.331*** (0.054)
Experiment	1 and 2	1 and 2	1	1

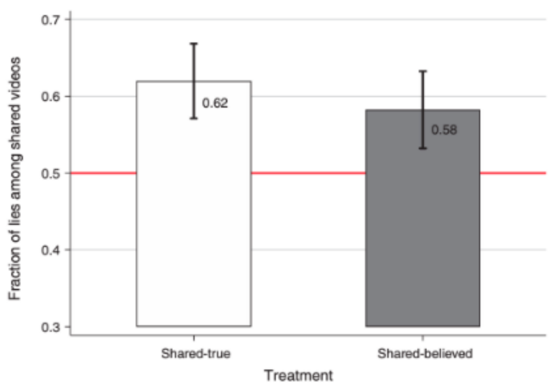
Note: Robust Standard errors are presented in parentheses.

Hypothesis 3:

The authors used the third experiment to answer this hypothesis. To do so, they built 2 treatments: the "Shared true treatment" and the "shared believed treatment" (explained in more detail in the **Summary of the experiment Experiments section**). The big difference between these 2 treatments is the Receiver 1s' incentives. In the first one, Receiver 1s earned a bonus if their video was watched by Receiver 2s and the video was true, and in the second one, Receiver 1s earned a bonus if the video was believed by Receiver 2s. The results obtained allowed the authors to analyze if the receivers sent more true videos when they were incentivized to do so.

From Hypothesis 1 and 2, we know that Receiver 1s ability to detect lies is not better than chance and they are significantly overconfident. In Experiment 3, when they were incentivized to share a truthful video, 62% of Receiver 1s shared lies in the Shared-true treatment, and 58.2% of them shared lies in the Shared-believed treatment. These numbers are reported on FIGURE 1 opposite, which comes from the paper (Sierra-Garcia & Gneezy, 2021), and are significantly better than 50%, which means that the fact that Receiver 1s share lies can not be attributable to pure chance. However, there is no statistical difference between the two treatments (p-value = 0.2923). In both cases, Receivers 1 are incentivized to share truthful videos but they still share lies, which confirms the statement made for Hypothesis 3.

FIGURE 1: *Sharing Rates of Lies for the Shared-true and Shared-believed treatments*



The authors also try to identify whether Receiver 1s inability to share true videos might be linked to their overconfidence. Therefore, they are testing overconfidence as the difference between Receiver 1s belief about their ability and their actual ability, and they are measuring it for Receiver 1s whose actual ability is below median, above median, and for all of them. The marginal effects on the likelihood that Receiver 1s share a lie are reported in TABLE 2.

We see that Receiver 1s are systematically overconfident, even if their true ability is below 50%, and therefore share more lies. In the last column of the table, the authors include the variable "ability above median". We see that ability above median has a negative effect on the likelihood that Receiver 1s share a lie, however this effect is not statistically significant. Thus, a person that is better than the median to detect lies will have the same probability to share fake news as someone whose ability is below the median. As a conclusion, we can see that even when they are incentivized to share truthful information, Receiver 1s systematically share lies.

TABLE 2: *Who among overconfident receivers, or receivers whose ability is above median are more likely to share a lie?*

	Share Lie = 1		
	Ability below median	Ability above median	All
Overconfidence (belief about ability - ability)	0.283*** (0.102)	0.456*** (0.113)	0.402*** (0.084)
Ability above median			-0.067 (0.051)

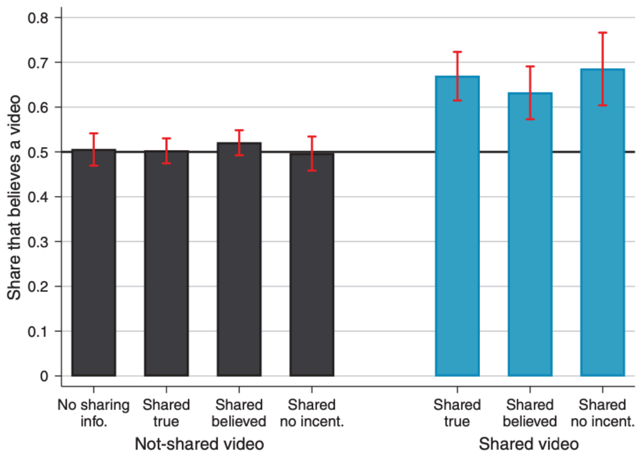
Note: Robust Standard errors are presented in parentheses.

Hypothesis 4:

Through this hypothesis, the authors wanted to analyze the impact of sharing a video on others beliefs. The authors tested this hypothesis in the last experiment. To do so, they added 2 treatments, the Shared-no-incentive treatment and the No-sharing information treatment. Receiver 1s were not paid any bonuses if Receiver 2s' guesses were correct in the no-sharing information treatment. The authors could then compare these results with the results obtained in the first 2 treatments (in which Receivers 2 knew which videos had been shared by Receiver 1s). This also allowed the author to analyze whether the belief of Receivers 2 was impacted by the incentive given to Receiver 1s.

FIGURE 2 below, which comes from the paper (Sierra-Garcia & Gneezy, 2021) shows that Receiver 2s chose to watch and believe videos shared by Receiver 1s in 75.5%, 69.5% and 66.1% of the cases in the Shared-true, Shared-believed and Shared-no-incentive treatments respectively, whereas these percentages equal approximately 50% when the videos are not shared.

FIGURE 2: *Comparison of the percentage of videos watched and believed by Receiver 2s when the videos are not shared versus when they are shared*



The following table (TABLE 3) shows coefficients obtained by regressing OLS on the likelihood that a video is watched (columns (1) to (3)) and believed (columns (4) to (6)). The first three columns are independent from the last three columns. The overall effect of shared videos on the likelihood that a video is shared and on the likelihood that a video is believed is captured by the variable "Shared video"; all the coefficients are positive and statistically significant. Therefore, it shows that a video that is shared is more likely to be watched and believed. Column (6) contains video fixed effects which indicate that, conditional on watching the same video, Receiver 2s are 8% more likely to believe a video when it is shared. Even though sharing a video increases the likelihood for a video to be watched, the effects are weaker for the Shared-believed (-0.068*) and the Shared-no-incentive (-0.107***). Indeed, the coefficients are negative (and statistically significant). Videos shared in the two treatments are also less likely to be believed (coefficients in columns (4) to (6) are negative), though the results are not statistically significant.

TABLE 3: *Coefficients of the OLS regression on the likelihood that a video is watched and believed, using a Linear Probability Model*

	Choose video to watch = 1			Believe video = 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Shared video	0.291*** (0.024)	0.293*** (0.029)	0.268*** (0.024)	0.167*** (0.031)	0.156*** (0.038)	0.085*** (0.030)
Shared-true	-0.036** (0.016)	-0.037** (0.016)	-0.033** (0.016)	-0.003 (0.023)	-0.003 (0.023)	0.004 (0.022)
Shared-believed	-0.028* (0.016)	-0.028* (0.016)	-0.025 (0.016)	0.015 (0.023)	(0.015) (0.023)	0.026 (0.022)
Shared-no-incentiv e	-0.022 (0.019)	-0.022 (0.019)	-0.019 (0.018)	-0.009 (0.027)	-0.010 (0.027)	-0.011 (0.025)
Shared-believed x shared video	-0.068* (0.035)	-0.070** (0.035)	-0.064* (0.035)	-0.055 (0.046)	-0.055 (0.046)	-0.063 (0.044)
Shared-no-incentiv e x shared video	-0.107*** (0.044)	-0.113*** (0.044)	-0.108*** (0.044)	0.022 (0.055)	0.023 (0.055)	0.046 (0.054)
Vidéo fixed effects	No	No	Yes	No	No	Yes

Note: Robust Standard errors are presented in parentheses.

Motivating Receiver 1s to share truthful videos would increase the believability of the latter. As shown in column (4), for the "Shared video" variable, a shared video had 16% more chance to be believed in the Shared-true and the Shared-no-incentive treatments. A reason for this could be that in the Shared-true treatment, Receiver 1s were incentivized to share a truthful video, and Receiver 2s knew those incentives; therefore, Receiver 2s' opinion about the veracity of the videos could have been influenced, because they knew that the Receiver 1s would earn a bonus if their guess was right. Receiver 2s knew that it was not in the interest of the Receiver 1s to share lies, since they would not earn any bonus in that case. In the Shared-no incentive treatment, the authors removed the link between Receiver 1s' bonus payment and Receiver 2s' choice of the video. Receiver 1s were only told to share a video they found interesting. This characteristic of the treatment might have influenced Receiver 2s belief, since there was no chance they might be influenced to have the same opinion as Receiver 1s (as in the Shared-true treatment, where the videos were supposed to be true). Also, the fact that Receiver 1s would not earn any bonus could have put less pressure on Receiver 2s beliefs, and therefore they would have been more likely to believe the videos shared . These findings confirm Hypothesis 4.

However, a video shared in the Shared-believed case had less chance to be watched (the coefficients are negative and statistically significant), but has equal chance to be believed since the coefficients are not statistically significant. These results show that, in this treatment, Receiver 2s belief might have been influenced by the fact that they knew Receiver 1s were sending a video they thought Receiver 2s might believe. Receiver 2s might therefore have anticipated the fact that Receiver 1s intentions would not necessarily be to share a truthful video, and that their motivation to share would have mostly been driven by the monetary incentive. Receiver 2s would therefore have shown more skepticism. The results obtained for the videos shared in the Shared-believed treatment constitute a limiting condition for the effect of sharing, and do not confirm Hypothesis 4.

It is also interesting to look at the percentage of Type I errors made by Receiver 2s (i.e. when Receiver 2s assess that the video they watch is true when, in fact, it is false) in the case the videos are not shared compared to the one when they are shared.

The authors observed that when the videos are not shared, the percentage of Type I errors are similar for all treatments (27% on average) and are not significantly different from 25%. However, when the sharing effect is added, Receiver 2s believe lies significantly more (43.1%, 34.1% and 44.9% in the Shared-true, Shared-believed and Shared-no-incentive respectively). Again, we see that the number of Type I errors are higher in the Shared-true and Shared-no-incentive treatments, and this tendency is weaker in the Shared-believed treatment, as discussed above.

The No-sharing-information treatment can be considered as the control group for experiment 3. The videos shown to Receiver 2s in the context of this treatment were not shared, and therefore no monetary incentives were provided. It is interesting to say that there is no significant difference between the three treatment groups and the control group in the case the videos are not shared, even if Receiver 2s are told that the videos they watch are true (Shared-true treatment) or have a great chance to be believed (Shared-believed treatment). Nevertheless, when the sharing effect appears, there is a significant increase in the percentage of type I errors. Hypothesis 4 is then verified.

Critical analysis

The paper has some potential problems in the identification strategy which could question the results that the researchers found. First of all, the number of people participating varies significantly depending on the experiment, it ranges from 185 to 1056 receivers and 10 to 42 senders per experiment. Since the variance for the smaller samples will be higher than the variance for the bigger samples, this might cause a problem. For example, we saw that Experiments 1 and 2 found different results for Hypothesis 1, even though the two experiments were identical. If the sample sizes of experiment 1 (380 receivers and 10 senders (no-prior), 192 receivers and 10 senders (prior)) were bigger and closer to the sample size of experiment 2 (1056 receivers and 42 senders), the variance in the results of experiment 1 would have probably been smaller and closer to the variance in the results of experiment 2. The results for hypothesis 1 would have been more precise and the experimenters could have more easily confirmed the veracity of hypothesis 1. One way to improve this work would therefore be to homogenize the sample sizes (don't have to be exactly similar) so that the difference in results is only due to the difference in treatment between experiments.

Internal validity problems

Concerning the internal validity, there is an experimental effect problem. The situation at which the receiver finds himself is very different from an actual usual situation (being in a lab experiment and getting pay bonus incentives VS scrolling through social media on the phone before going to sleep for example). This could cause the results to be biased. For example, the results for hypothesis 1 could be positively biased, because the receivers pay more attention because they are in an experiment and are getting paid, so they get better at detecting lies. Maybe the results for hypothesis 1 could be negatively biased instead, because the receivers could be more stressed in a lab situation and therefore get worse at detecting lies. Unfortunately, this experimental effect is something that can't really be solved. As Levitt and List (2005) have pointed out in a research paper that studies the experimental effect in general: "lab experiments are a useful tool for generating qualitative insights, but are not well-suited for obtaining deep structural parameter estimates." Because lab experiments have an experimental effect, the results can't fully represent the real situation.

External validity problems

Concerning the external validity, there is a non-representative sample problem. All the receivers are 18 years old or older, which is not representative of the actual population in social media. For example, the results of Hypothesis 1 are probably positively biased. Indeed, the sample of receivers doesn't include teenagers (17 years old or less) who might have a worse ability to detect lies than if they just gave an answer randomly, since teenagers can be more easily influenced (so they might be even more overconfident in their ability to detect lies too. So in this case, if teenagers were included in the sample, the results for hypothesis 2 could be positively biased.).

These variance and bias issues could end up being an issue for the findings of this paper. Since a small change in the numerical results for some of the experiments could slightly change the findings of an hypothesis (example: for hypothesis 1), it's important that the results have the least bias and variance possible.

Strength in the randomization process for the receivers

Since the experiment follows a randomized control trial (RCT) strategy, it also has some significant strengths. Thanks to the randomization process, most characteristics of the individuals are correctly represented. For example, the proportion of men and women is 50%-50%, which approximately corresponds to the actual population, so there isn't any bias on this part. Thanks to the fact that the most characteristics of the individuals are correctly represented, the number of omitted variables is heavily reduced (if there are any).

The experimenters also made sure that there was no experimenter bias when they chose the news events for the senders. They did that by choosing only news events dating exactly 6 months prior to when the videos were recorded. Thus the experimenters were not able to choose the news events based on their preferences.

Additional Analysis

We noticed that the sample of our paper was very broad. Indeed, the participants of the experiment were between 18 and 84 years old. We wondered whether the ability to detect the truthfulness of a video was the same across all age groups. Our basic assumption was that it would decline as the age of the receivers increases. We therefore decided to focus our additional analysis on this question. To do so, we put forward two hypotheses.

Hypothesis 1:

"Older people tend to be worse at detecting lies than younger people".

This hypothesis is based on the fact that we believe that older people are less used to being confronted with fake news and therefore have less ability to detect them than younger people. Indeed, although fake news has always existed, it has developed enormously with social media. The older generations, born before the internet was invented, were less confronted with fake news on a daily basis and trusted more the newspapers or TV. On the contrary, young people tend to trust classical media less and to be much more critical of the news, even if they are considered as reliable sources. Interestingly, another research paper entitled *Aging in Era of Fake news* (Nadia M. Brashier and Daniel L. Schacter, 19 May 2022) found similar results when it observed that older people were the portion of the population that shared the most fake news during the 2016 U.S. election.

We decided to separate the sample into 4 different age groups of the same range as the one used in the study to determine if the ability to detect lies differs by age groups. Our assumption was that as a person ages, his or her ability to detect lies decreases.

Hypothesis 2:

"Men are more overconfident than women but this overconfidence tends to decrease with age"

We decided to focus our second hypothesis on two fixed variables over which receivers have no power, namely age and gender. As we observed in our article, men tend to be more confident than women. However, we believe that men's overconfidence tends to decrease with age and move closer to women's. We based this hypothesis on the fact that young people often tend to be very self-confident, especially young men. However, as men age, they tend to be more mature and aware of their personal limits, thus less overconfident. We believe that although older men remain more overconfident than older women, this difference tends to decrease.

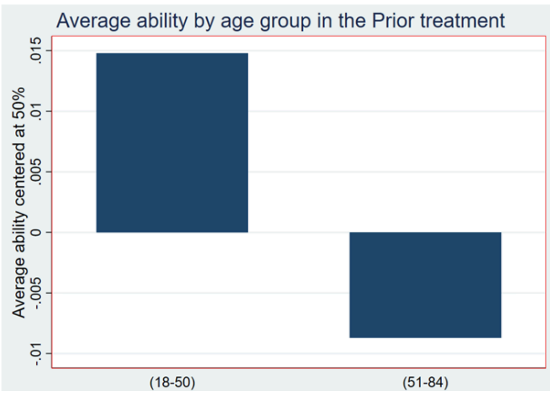
Results of the Additional Analysis

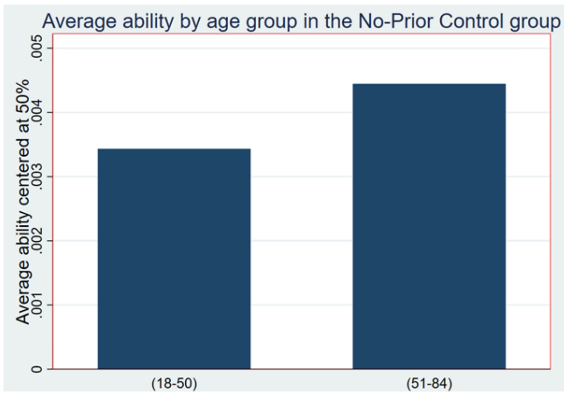
<u>Hypothesis 1:</u>

To isolate the effect of age we compared the prior (treatment) to the Non-Prior (Control group) with the data from experiment 1 as it was the only experiment where we had the Prior treatment. However, by separating the data into 4 groups we reduce the initial sample size by a lot. Indeed, as we can see in the first graph, there was only one participant in the prior group over 66 years old and only 7 participants in the No-Prior group. The results were therefore not reliable at all.

We therefore decided to extend our additional analysis by separating the population into 2 age groups (18-50 vs 51-84). This allowed us to see if a larger sample size would give a significant result. However, even when we separate the data into two groups, the two graphs show very small differences in the ability to detect lies. As we can see in the table below, results were not significant (p-value = 0.183 for Prior treatment and p-value=0.525 for Non Prior) which means our first hypothesis is either False or the insignificant t-test are due to small sample sizes in older age groups.

Groups	4 Groups				2 Groups	
Prior	36	115	40	1	169	23
No-Prior	120	182	71	7	335	45
Age	18-29	30-44	45-65	66-84	18-50	51-84
Prior Ha: Difference between means . P-values	18-29 vs 30-44	18-29 vs 45-65	18-29 vs 66-84		18-50 vs 51-84	
	0.227	0.139	.		P-value=0.183	
	30-44 vs 45-65	30-44 vs 66-84	45-65 vs 66-84			
	0.284	.	.			
No-Prior Ha: Difference between means P-values	18-29 vs 30-44	18-29 vs 45-65	18-29 vs 66-84		18-50 vs 51-84	
	0.326	0.448	0.652		P-value = 0.525	
	30-44 vs 45-65	30-44 vs 45-65	45-65 vs 66-84			
	0.593	0.703	0.648			

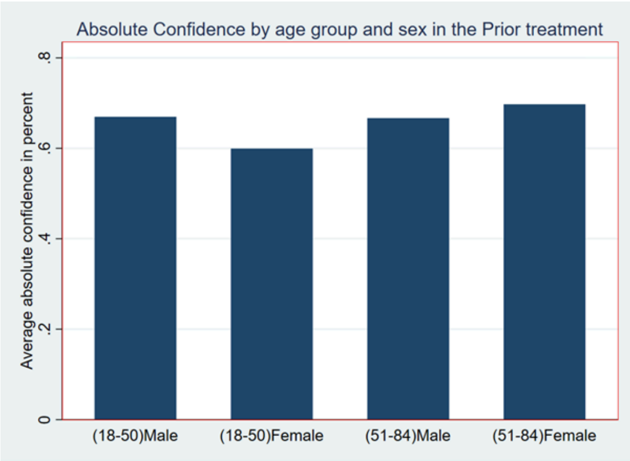


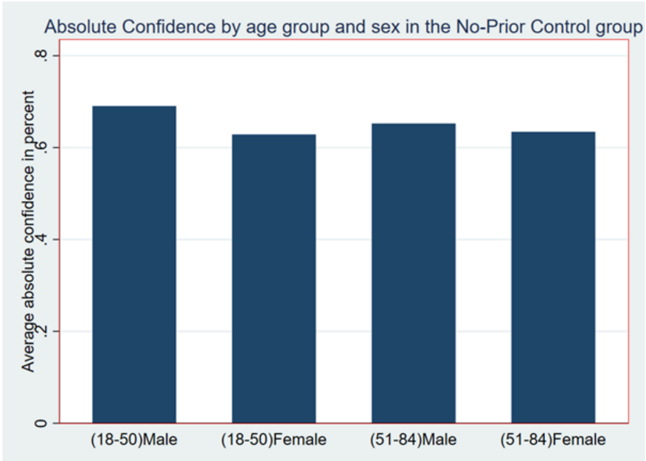


As you can see on the two graphs above, the difference between the two treatments are very small and therefore not relevant.

<u>Hypothesis 2:</u>

First we ran our hypothesis by comparing the Young-male and Young-female groups. If we observe the two graphics below, it looks like the gap between the genders is smaller in the Old-male vs Old-female groups than between Young-male vs Young-Female ones in Prior and No-prior, which would confirm our hypothesis.





In the tables 5 and 6 below, we distinguish 2 groups: group 0 represents male Receivers and group 1 represents female Receivers. The two tables show Two-sample t-tests, one for the No-prior treatment (graph on the left) and the other for the Prior treatment (graph on the right). By running t-tests, we were able to compare whether young men were more overconfident than women.

In the Non-Prior group, we found significant results for a 95% CI (p-value=0.0004), proving that Young-men are more overconfident than Young-females by 7 percentage points and 11 point in the Prior (p-value=0.0015). As the prior is the treatment it will mean that the effect of overconfidence by gender when young is due to age is of 4 (11-7) percentage point.

We found similar results in the No-prior group. When we compared the Old-Male vs Old- Female groups we found no evidence that their confidence was significantly different, which would have confirmed our hypothesis that overconfidence tends to decrease in older age categories. However we think that those results are (sadly) due to the small number of "old" people that took part in the experiments of the paper. Indeed, as we can see in the table opposite, the samples for older male and older female are less than 20 in the prior and less than 25 in the No-prior, which makes finding a significant result very unlikely.

TABLE 4: Sample sizes by treatments, age and gender

Groups	Young		Old	
Prior	87	82	6	17
No-Prior	197	138	23	22
Age	18-50	18-50	51-84	51-84
Gender	Male	Female	Male	Female

TABLE 5: Young-men vs Young-women in the No-Prior treatment

```
. ttest percentguess if experiment==1 & treatment==1, by (age_Gender_Young)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |     197    .6901015    .0113147    .1588093    .6677873    .7124157
       1 |     138    .6282609    .0144722    .1700104     .599643    .6568787
---------+--------------------------------------------------------------------
Combined |     335    .6646269    .0090743    .1660868    .6467769    .6824768
---------+--------------------------------------------------------------------
    diff |            .0618407    .0181508                 .026136    .0975453
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   3.4071
H0: diff = 0                                     Degrees of freedom =      333

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9996         Pr(|T| > |t|) = 0.0007          Pr(T > t) = 0.0004
```

TABLE 6: Young-men vs Young-women in the Prior treatment

```
. ttest percentguess if experiment==1 & treatment==2, by (age_Gender_Young)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      87    .6695402    .0164384    .1533267    .6368619    .7022186
       1 |      82    .5993902    .0164048    .1485514    .5667499    .6320306
---------+--------------------------------------------------------------------
Combined |     169     .635503    .0118947    .1546306    .6120207    .6589852
---------+--------------------------------------------------------------------
    diff |              .07015    .0232455                .0242571    .1160428
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   3.0178
H0: diff = 0                                     Degrees of freedom =      167

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9985         Pr(|T| > |t|) = 0.0029          Pr(T > t) = 0.0015
```

Conclusion

This paper allowed the authors to show the factors that lead someone to share fake news and how quickly they can spread.

The results of hypothesis 1 and 2 show that people are bad at detecting lies and are overconfident in their ability to do so. The results of hypothesis 3 and 4 show that when people are incentivized to share a truthful video, the videos they share are more likely to be watched and believed, but are also more likely to contain lies. The combination of these results with people's low ability to detect lies might explain why fake news tend to be spread and believed so easily

In our critical analysis, we have found that there is a variance issue, an experimental effect and a non-representative sample problem. These issues slightly bias the results and make them vary a little bit more. This might cause the findings for some hypotheses to be slightly different but these issues are not big enough to impact the general findings of the paper.

FInally, our additional analysis offers a better understanding about the age groups that are more likely to be affected by fake news. The results of hypothesis 1 shows that old people aren't worse at detecting lies than young people. Via hypothesis 2, we found that men are in general more confident than women, but that this tends to decrease with age. However, this result was not very reliable because the sample size was too small for the elderly. It would therefore be interesting to extend this paper to analyze whether this phenomenon is also observable for a larger sample.

References

Sierra-Garcia M. and Gneezy U. 2021 "Mistakes, Overconfidence and the Effect of Sharing on Detecting Lies." *American Economic Review 111(10): 3160-3183*

Levitt S. and Litt J. (2005) "What Do Laboratory Experiments Tell Us About the Real World?", viewed 8 december 2022, retrieved from https://www.researchgate.net/publication/248419698_What_Do_Laboratory_Experiments_Tell_Us_About_the_Real_World>

Nadia M. and Daniel L. (2022) *"Aging in Era of Fake news"*. National Center for Biotechnology information. From: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7505057/