

## **People are worse at detecting fake news in their foreign language**

Rafał Muda<sup>1</sup>, Gordon Pennycook<sup>2</sup>, Damian Hamerski<sup>1</sup>, and Michał Białek<sup>3</sup>

*1 Faculty of Economics, Maria Curie-Skłodowska University, Poland*

*2 Department of Psychology, University of Regina, Canada*

*3 Institute of Psychology, University of Wrocław, Poland*

# **People are worse at detecting fake news in their foreign language**

## **ABSTRACT**

Across two preregistered within-subject experiments (N = 570), we found that when using their foreign language, proficient bilinguals discerned true from false news less accurately. This was the case for international news (Study 1) and for more local news (Study 2). The difficulty in discerning the two types of news occurred, because when using a foreign vs native language, true news headlines was judged as equally (Study 2) or less believable (Study 1) while false news was judged as more believable (Studies 1 and 2). Contrary to past theorizing, we observed the foreign language effect was neither mediated by perceived arousal of news, (Study 1), nor by individual differences in cognitive reflection, (Studies 1 and 2). Finally, using signal detection theory modelling, we showed that the negative effects of using a foreign language were not caused by adopting different responding strategy (e.g., preferring omissions to false alarms), but by decreased sensitivity to truth.

**Keywords:** fake news, bilingualism, foreign language effect, cognitive reflection, arousal

## **Wordcount**

Abstract: 153

Main text: 3944

## **People are worse at detecting fake news in their foreign language**

The spread of falsehoods on social media is emerging as a global challenge that may threaten democracies (Lee, 2019). The internet has provided unprecedented access to information; unfortunately, it has also increased the reach of misinformation. An element of this that has not received much attention is the fact that some people get their news in a foreign language (e.g., immigrants who need to navigate their new homelands or people who live in countries where the state controls most of the media and foreign sources are needed). This is a context, therefore, where misinformation and “fake news” (i.e., fabricated news headlines that are presented as if they are legitimate and spread on social media; Lazer et al., 2018) may be particularly pernicious. In this project, we empirically tested whether reading news headlines in a foreign language affects readers ability to discern true from fake news.

Suitably, there has been a surge of research devoted to misinformation and fake news. A key argument is that people tend to believe fake news more because they find it intuitive (Pennycook & Rand, 2019, 2020; Ross et al., 2019) or emotionally evocative (Martel et al., 2019). For example, Pennycook and Rand (2019) found that people who are dispositionally more reflective and less intuitive are better able to discern between true and false news content regardless of whether it is consistent or inconsistent with their political ideology. This research shows commonality between belief in fake news and other forms of epistemically suspect beliefs such as conspiratorial beliefs, delusional ideation or belief in alternative medicine (Pennycook & Rand, 2020) which have long been associated with an overreliance on intuition and a lack of reflective thought (see Pennycook et al., 2015 for a review).

Despite this convergence, a central question remains: If an overreliance on gut feelings and intuitions supports belief in fake news, what triggers people to stop and reflect on a given fake news headline? Understanding this key element is central to developing interventions to increase people’s capacity to detect falsehoods on social media. This uncertainty applies both to understanding fake news but also other forms of epistemically suspect beliefs and their association with reflective thought.

A potential solution can come from the research on the foreign language effect (FLe), which is reported to help bilinguals make better, less biased decisions in their foreign language (Costa, Foucart, Arnon, et al., 2014; Keysar et al., 2012). Although the FLe is

thought to benefit decision-making, its exact mechanism remains an open question. Several proposed mechanisms include lower emotional engagement, greater cognitive engagement, or more abstract construals (Hayakawa et al., 2016); in each case, the underlying mechanism is that thinking in a foreign language helps bilinguals to use their reflection to override gut feelings. Specifically, biased decisions are often rooted in intuitively appealing, affect rich responses which win against reason based, more abstract responses (De Neys & Pennycook, 2019; Evans & Stanovich, 2013; Pennycook et al., 2015a, 2015b). If a person's intuitions are less vivid, or reflection is more eagerly triggered, reasoners should engage in more reflection and, in turn, increase their consistency and normative accuracy. This suggests that people using their second language should be better able to discern between true and false news content than when using their native language.

More recent research found that decision-makers do not always benefit from using their second language: using foreign language does not affect decision-making in gambling (Muda, Walker, et al., 2020), intertemporal choice (Białek et al., 2022; Xu & Liang, 2021), and may not actually increase reflective thinking (Mækelæ & Pfuhl, 2019). Moreover, it can distort metacognition, i.e., the ability to detect when a salient intuition is wrong (Białek et al., 2020). Indeed, when information is presented to bilinguals in their foreign language, related brain potentials suggest more effort is required to compare new information to prior knowledge (Romero-Rivas et al., 2017), possibly muting the intuitive believability judgments of news. In the light of these findings, using a foreign language should have no effects, or even result in higher susceptibility to fake news and decreased ability to discern between true and false news content compared to using native language<sup>1</sup>. We tested this prediction experimentally.

## Study 1

To test the effects of using a foreign language on the ability to discern true from false news, we designed an experiment in which bilingual participants were asked to judge their believability of news headlines. Half of the news was false. The manipulation was that half of true and half of false news was randomly selected to be displayed in a foreign language. The

---

<sup>1</sup> A recent study reported null effects of language on credibility of fake news (Fernández-López & Perea, 2020). However, their design included false news only, not allowing to infer the effects of using a foreign language on the ability to discern true from false news.

goal of the experiment was to compare (within-subject) the ability to discern true from false news presented to bilingual participants in their native language (Polish) vs. in their foreign language (English).

We preregistered our experiments: <https://aspredicted.org/blind.php?x=dv2hn9> (Experiment 1), and <https://aspredicted.org/ni2sd.pdf> (Experiment 2). Data, code, and materials are to be found: <https://osf.io/km4eu/>. Data was analyzed using JAMOVİ 2.2.5 (The Jamovi Project, 2021). We also report an earlier experiment in the supplementary materials that did not offer a good test of the hypothesis because the true and false news that were used were indistinguishable from each other in terms of their believability.

## Method

### Participants

We analyzed the data from  $N = 300$  participants (188 females,  $M_{\text{Age}} = 22$ ,  $SD = 3.4$ ;  $N = 177$  native language condition;  $N = 123$  foreign language condition). Participants were recruited through an external research firm. Participants were deemed ineligible prior to entering the survey if they declared that Polish is not their native language, had a parent that speaks English as their native language, lived more than 10 months in an English speaking country, or self-rated their global proficiency in English as 4 or lower on a scale from 1 to 10. This selection criteria was preregistered and roughly follows selection criteria in past work; Białek, Muda, et al. (2020); Costa, Foucart, Arnon, et al. (2014); Keysar et al. (2012); Muda, Pieńkosz, et al. (2020).

We initially recruited 301 participants but excluded 1 participant due to failed translation task that served as a proficiency verification. Subjects received 20 PLN (about 5 USD) for participation in 15 minutes' laboratory experiment. Our sample size provided us with  $\beta = .80$  power to detect correlations as small as  $r > .14$ , and differences in two within-subject conditions of  $d > .14$  (Faul et al., 2009).

### Materials and Procedure

Participants evaluated fake news in two language conditions with counterbalanced order: native language (L1: Polish) vs. foreign language (L2: English). Their task was to read 16 news headlines (8 true and 8 fake, see Supplementary Materials for exact wording) and to judge their believability (1 – Extremely unbelievable; 7 – Extremely believable), willingness

to share the news (I never share anything online; No; Maybe; Yes), familiarity with the news (No; Unsure; Yes), and how emotionally arousing is the news (1 – Not at all; 7 – Extremely). Except for using a foreign language in some news, the procedure regarding the selection and presentation of fake news closely mirrored previous studies (Pennycook, Binnendyk, et al., 2021). The fake news headlines were selected from Snopes.com, independent fact-checking website. The real headlines were contemporary stories from mainstream news outlets. News was presented as screenshots in the same format as if they were shared on Facebook, albeit with manipulation (i.e., language of the original news was changed but maintaining the design of the news, pretending to be from the BBC). News were translated to Polish, and back-translated by the bilingual authors of the study. For each participant, we randomly selected four headlines to be presented in their native language and remaining four headlines to be presented in a foreign language. The order of presentation of languages was counterbalanced.

Between the two language trials, participants completed a filler task (i.e., delay discounting), presented in the same language as the first trial. As their very last task, participants solved the Cognitive Reflection Test (CRT, Frederick, 2005), presented to them in their native language. CRT consists of three mathematical tasks that have a strongly appealing intuitive but incorrect responses. For example, participants are asked how much a ball costs, given bat and ball cost together \$1.10 and the bat is \$1 more expensive than the ball. CRT consists of three such items. Sum of correct responses serves as a measure of cognitive reflection. No other data was collected.

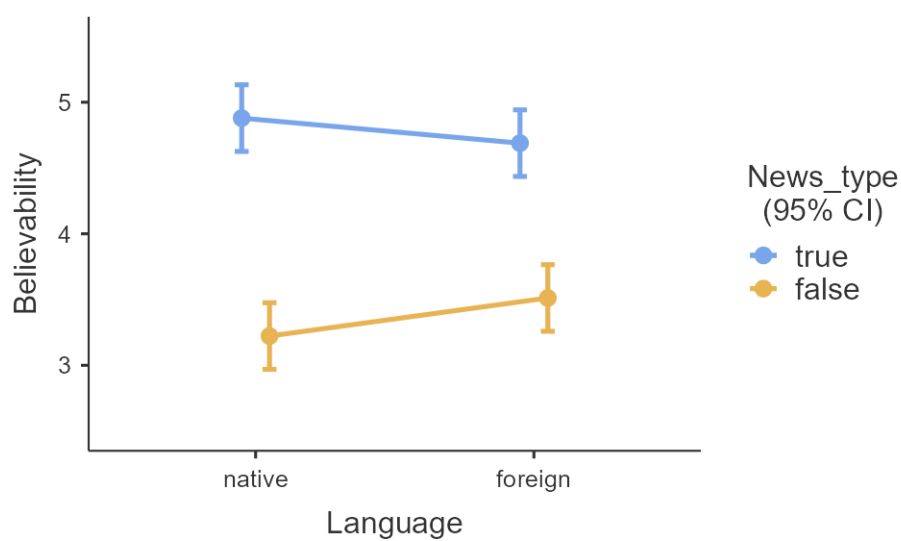
## **Results**

Our preregistered analysis tested the effects of using foreign language on the believability of true and false news by controlling individual differences in CRT. Our main interest was in the main effect of news type informing us on whether people were able to discern true from false news, and news type by language interaction, informing us whether this ability was increased or decreased in L2 users. In the model, news\_ID and participants\_ID were clustering variables; type of news (true/false) and language (L1/ L2) were factors; CRT score and arousal were covariates. We contrast-coded our factorial predictors as -0.5 and +0.5, and mean-centered the continuous predictors.

The results are presented in Table 1. False news were judged as less believable, but this difference was larger in L1, ( $\Delta_{\text{bel}} = -1.55 [-1.90, -1.21]$ ,  $t(16.1) = -9.55$ ,  $p < .001$ ) than in

L2 ( $\Delta_{\text{bel}} = -1.11 [-1.56, -0.77]$ ,  $t(16.1) = -6.85$ ,  $p < .001$ ), see Figure 1, panel A. This decreased ability to differentiate true from false news in L2 was driven by both, increased believability of false news, ( $\Delta_{\text{bel}} = 0.28 [16, 39]$ ,  $t(4481) = 4.59$ ,  $p < .001$ ) and decreased believability of true news ( $\Delta_{\text{bel}} = -0.16 [-0.28, -0.05]$ ,  $t(4480) = -2.72$ ,  $p = .007$ ). We also observed that cognitive reflection (regardless of language used) increased believability of true news,  $B = 0.14 [0.05, 0.23]$ ,  $t(440) = 3.12$ ,  $p = .002$ , but had no impact on the believability of false news,  $B = 0.00 [-0.09, 0.09]$ ,  $t(440) = 0.05$ ,  $p = .960$ , see Figure 1, panel B.

A



B

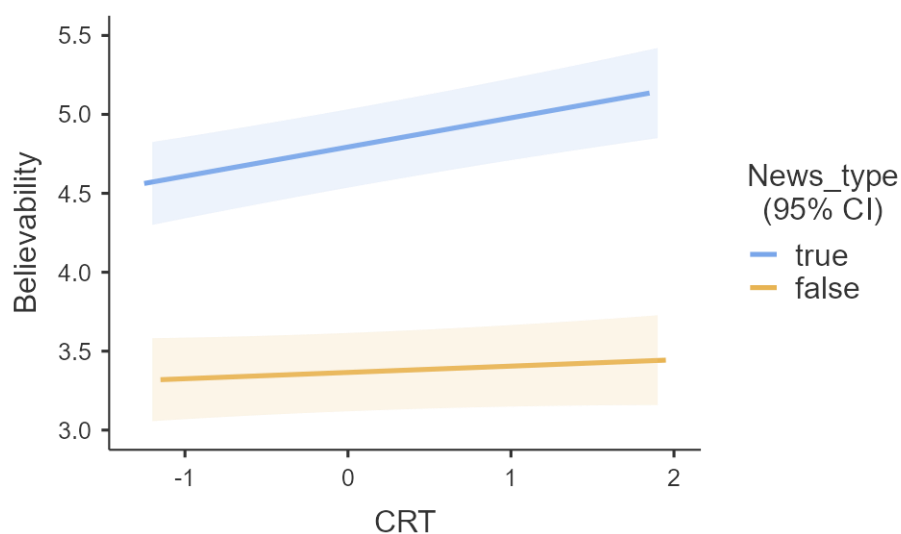


Figure 1.

Effects of language and news type on believability ratings. Panel A presents the effects of language, and Panel B of cognitive reflection (standardized) on believability ratings of news.

We followed-up the preregistered analysis by additionally including perceived arousal of the news as a covariate, with all its first and second level interactions. This analysis has two benefits: it parses out part of the variance in believability ratings, allowing for better test of the unique contribution of the FLe. Additionally, by including the language interaction with arousal and with cognitive reflection, we can test whether the FLe is moderated by arousal of news (i.e., whether it works only in arousing stimuli) and by cognitive reflection (i.e., affects mostly people who typically rely on intuition when thinking).

The previously reported results are unaffected such that the effects of foreign language nor the language by news type interaction did not interact with CRT or with perceived arousal (Table 1). There was a main effect of perceived arousal, suggesting that more arousing news were perceived as more believable. CRT scores and perceived arousal also interacted, such that the effects of arousal on believability decreased as a function of the CRT. Finally, using L2 did not affect neither the CRT performance, nor the perceived arousal of news.

Table 1. Results of the Linear Mixed Effects analysis

	<b>Model 1</b>		<b>Model 2</b>	
	<i>B</i>	<i>p</i>	<i>B</i>	<i>p</i>
Intercept	4.08	< .001	4.07	< .001
CRT	0.07	0.091	0.11	0.002
News type	-1.33	< .001	-1.42	< .001
Language	0.06	0.185	0.05	0.234
News type x Language	0.44	< .001	0.48	< .001
CRT x News type	-0.14	< .001	-0.14	< .001
CRT x Language	0.02	0.576	0.01	0.694
CRT x News type x Language	-0.06	0.417	-0.07	0.332
Arousal	-	-	0.26	< .001
News type x Arousal	-	-	0.03	0.200
Language x Arousal	-	-	0.00	0.940
CRT x Arousal	-	-	-0.03	0.020
News type x Language x Arousal	-	-	< .001	> 0.999
CRT x News type x Arousal	-	-	-0.02	0.467



CRT x Language x Arousal	-	-	0.02	0.295
<b>Random Components</b>				
	SD	ICC	SD	ICC
ID	0.70	0.18	0.58	0.14
News	0.30	0.04	0.31	0.04
<b>Model</b>				
Observations	4800		4800	
N ID	300		300	
N News	16		16	
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.147 / 0.328		0.222 / 0.359	

## Study 2

Study 1 indicates that people are worse at discerning between true and false news in their foreign language. In Study 2, we address the possibility that the results could be attributed (more mundanely) to foreign language causing decreased confidence in responding (Geipel et al., 2015a; Montero-Melis et al., 2020) or, simply, a more conservative responding strategy that results in more omissions. To this end, instead of having participants judge believability using a continuous scale, we asked participants to make yes/no classifications followed up by confidence ratings. Finally, we used more locally-based news rather than US-based news as in the Experiment 1.

## Method

### Participants

Data from N = 270 participants was analyzed (140 females, 130 males,  $M_{\text{Age}} = 25.1$ ,  $SD = 8.27$ , N = 141 native language condition; N = 129 foreign language condition ). Participants were recruited via Prolific Academic. We initially recruited 274 participants but excluded 4 participants due to failed translation task that served as a proficiency verification (i.e., they were asked to translate the text of the instruction). Compared to Study 1, we included additional exclusion criterion, i.e. self-rate level of stimuli understanding, but no one was excluded due to this criterion.

### Materials and Procedure

The procedure regarding the selection and presentation of fake news virtually mirrored Study 1. Participants evaluated fake news in one language condition: native language (Polish) or foreign language (English). Their task was to read 16 news (8 true and 8 fake, exact stimuli can be found: <https://osf.io/km4eu/>) and to judge their believability (YES/NO) and 3-point confidence ratings (Not at all, Moderately, Very), willingness to share the news (0 – 100 Likert scale (“not at all” to “very”). The news headlines were contemporary stories from Polish mainstream news outlets highlighting both international and Polish current issues. Fake news were found on <https://fakenews.pl/>; true news were found on <https://www.polsatnews.pl/> and <https://tvn24.pl/>. All news were presented in random order As a last task, participants also solved the CRT. No other data was collected.

## Results

We preregistered a linear mixed model (LME) that controls for item-level effects (i.e., variability of believability of given news, or item-specific FLe). Additionally, as recommended by Batailler et al., (2022) we preregistered the receiver operating characteristic analysis (ROC) to dissociate sensitivity (ability to differentiate true from fake news) from response bias (strategy to prefer type 1 or type 2 errors). By analogy to the research incorporating ROC's to study belief bias (Białek et al., 2020; Dube et al., 2010; Trippas et al., 2014), instead of simply using yes/no responses we also collected participants' confidence ratings (1 to 3), allowing for more precise estimation of the parameters.

### Confirmatory analyses

#### *LME analysis*

As in Study 1, our preregistered analysis tested the effects of using foreign language on the believability of true and false news by controlling individual differences in cognitive reflection. We contrast-coded our factorial predictors as -0.5 and +0.5, and mean-centered the continuous predictors. Clustering variable were participant ID's, and news number (from 1 to 16). The results are presented in Table 2.

Table 2.

Results of linear mixed models conducted for Experiment 2

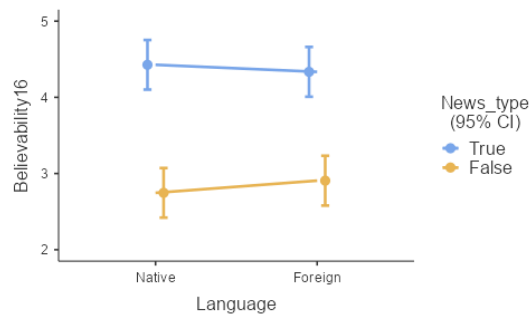
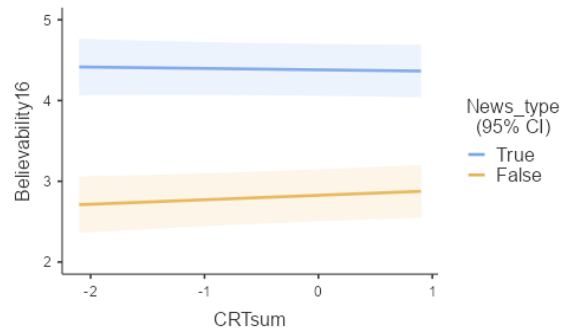
Believability	Confidence	Intent to Share
---------------	------------	-----------------

	<i>B</i>	<i>p</i>	<i>B</i>	<i>p</i>	<i>B</i>	<i>p</i>
Intercept	3.60	< .001	2.31	< .001	20.11	< .001
Language	0.03	0.529	-0.10	< .001	4.69	< .001
News type	1.55	< .001	-0.06	0.337	0.53	0.812
CRT	0.02	0.502	-0.02	0.035	-2.78	< .001
News type x Language	-0.25	0.014	-0.01	0.848	1.48	0.287
CRT x News type	-0.07	0.123	0.01	0.507	-0.38	0.552
CRT x Language	-0.01	0.885	0.00	0.896	1.97	0.052
CRT x News type x Language	0.06	0.531	-0.04	0.260	-1.25	0.327
<b>Random Components</b>						
	SD	ICC	SD	ICC	SD	ICC
ID	0.37	0.05	0.17	0.08	14.81	0.30
News	0.40	0.06	0.11	0.04	4.19	0.03
<b>Model</b>						
Observations	4320		4320		4320	
N ID	159		159		159	
N News	16		16		16	
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.166 / 0.246		0.010 / 0.116		0.0223 / 0.330	

As expected, we see that false news were judged as less believable than true news. Yet, this difference between the believability of true vs false news was larger in L1, ( $\Delta_{\text{bel}} = 1.68 [1.23, 2.13]$ ,  $t(15.6) = 7.86$ ,  $p < .001$ ) than in L2 ( $\Delta_{\text{bel}} = 1.43 [0.97, 1.88]$ ,  $t(15.9) = 6.65$ ,  $p < .001$ ), see Figure 2, panel A. This decreased ability to differentiate true from false news in a L2 was driven by increased believability of false news, ( $\Delta_{\text{bel}} = 0.16 [0.01, 0.31]$ ,  $t(4257) = 2.16$ ,  $p = .031$ ). In contrast to Study 1, the believability of true news in L2 decreased non-significantly ( $\Delta_{\text{bel}} = -0.09 [-0.24, 0.05]$ ,  $t(4257) = -1.25$ ,  $p = .213$ )<sup>2</sup>.

In contrast to prior research and to Study 1, CRT did not produce a main effect, nor interacted with news type. There was also no three-way interaction, when additionally controlling for the language in which the news was presented.

<sup>2</sup> As suggested by reviewers, we also explored whether the effect can be explained by low proficiency of our participants: the decreased ability to discern true from false news was driven by subsample of L2 participants with low proficiency who simply had comprehension issues. We replaced CRT scores with proficiency, leaving all other element of the analysis the same. The results showed no main effect of proficiency, no proficiency by news type interaction (with  $p = .075$ ), or news type by language by proficiency interaction. Hence, we believe our results are fairly robust to proficiency issues, but admit the power of our analysis could be fairly low. All analyses are reported in supplementary Materials.

**A****B****Figure 2.**

Effects of language and news type on believability ratings. Panel A presents the effects of language, and Panel B of cognitive reflection (standardized) on believability ratings of news headlines.

### *Confidence*

Geipel and colleagues (2015a) found that moral FLe was partially mediated by decreased confidence ratings. If using L2 could have simply reduced confidence, participants using it would report smaller differences in believability ratings between true and fake news; e.g., instead of false and true being correctly classified as “1” and “6” with the difference of “5”; one could classify the same news as “2” and “5” with the difference of “3”. In that case, people still classify news accurately as true or false, but only appear as less accurate because lower confidence in their classifications decreases the difference in continuous believability score. To see if this concern applies to our findings, we ran an LME with confidence as dependent variable, and all other predictors identical to the previous analysis (Table 2). The results showed that, indeed, confidence was lower in L2, but in contrast to Geipel and colleagues (2015a), it was not correlated to proficiency,  $r(2063) = .02$ .

Lower confidence in L2 could produce similar effects as a genuine difference in the ability to distinguish between true and false news. However, the reversed causal order between believability ratings and confidence can be true: participants could realize (at some level) that their classification of news in L2 is less accurate, and thus report lower confidence. We tried to address this problem using Signal Detection Theory (SDT; Green & Swets, 1966; Wixted, 2020).

### *Signal Detection Theory analysis*

If decreased difference in believability ratings of true and false news is attributable merely to lower confidence, we will see no effects of language on sensitivity (i.e., the ability to discern true from fake news), but we should see participants adopting more conservative response criterion (i.e., preference to rather classify true news as false than false news as true). Classification accuracy and the response bias can be estimated using ROCToolbox, (Koen et al., 2017). It fits all responses of all participants in a given condition to a ROC curve. Because the estimates for ROC are not accompanied by any measure of standard error, the comparisons between conditions are made descriptively.

A G-test indicated that the DPSD model fit both the L1 and L2 data well, both  $G's \geq 9.02$ ,  $p's \leq .029$  (Fig. 3), with the ROC curves  $adjR^2_{L1} = 0.958$ , and  $adjR^2_{L2} = 0.987$ . Points on the curve represent hit rates and false alarms, plotted cumulatively for each confidence level (Figure 3). Point 1 of the curve corresponds to responses “true” made with highest confidence; point 2 to responses “true” made with highest confidence and with moderate confidence; point 3 to responses “true” with all levels of confidence; point 4 to response “false” with lowest confidence and “true” with all confidence levels, point 5 to “false” with low and moderate confidence and all “true” responses with any confidence levels. Point 6 is always in the top-right corner.

The accuracy is visualized by the size of area under the curve (AUC). The larger the AUC, the greater the estimated accuracy. With perfect responding, hits = 1, and false alarms = 0, the points on the curve are in the top-left corner, and the  $AUC = 1$ . Diagonal represents a “guessing line”, where hits are just as frequent as false alarms; here the  $AUC = .5$ . In case the AUC is smaller than .5, participants simply responded incorrectly more often than correctly, which is typically attributable to not following the instruction properly. The response bias is visualized by the positioning of the points on the curve: For example, points closer to the top-

right corner include more hits and more false alarms: a liberal responding strategy.

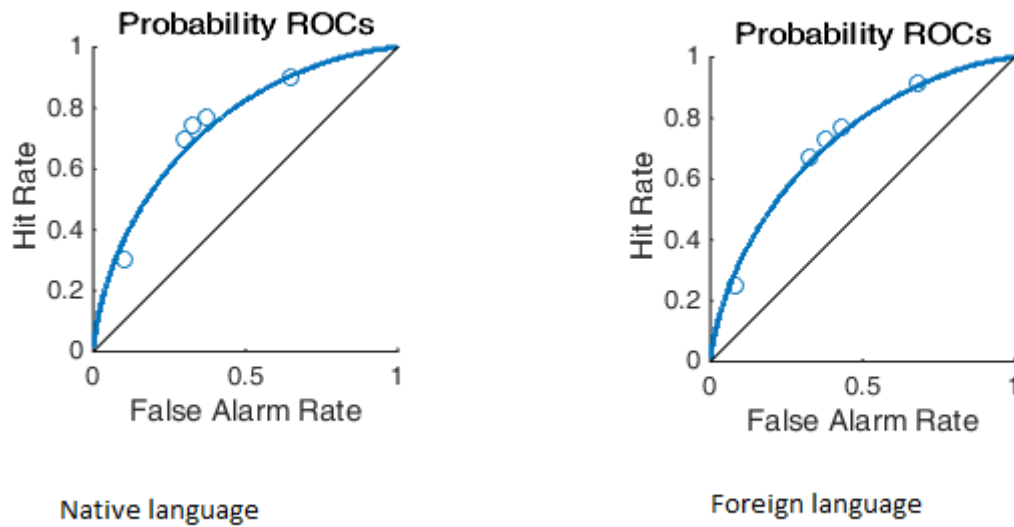


Figure 3.

Receiver operating characteristics (ROC) curves for discriminating between true and false news headlines in native vs foreign language

One of the simplest measures of the accuracy in ROC's is captured by hit rates minus false alarm rate. Since both are within a 0-1 range, perfectly accurate participant would make 100% hits, and 0% false alarms, thus this parameter would be 1.0. A guessing participant would have similar values of hits and false alarms; thus this parameter would be close to 0. In our data, the parameter was 0.41 for L1 condition participants, and 0.35 for L2 participants. Alternative measures of accuracy suggest similar conclusions:  $AUC_{L1} = 0.73$ , and  $AUC_{L2} = 0.71$ , and  $d'_{L1} = 1.09$ , and  $d'_{L2} = 0.91$ . So, alleviating our concerns about confounding lower confidence with decreased accuracy of classification, we see that accuracy decreased when headlines were presented in L2 across all metrics.

ROC analysis allows also to investigate whether participants in L2 adopted different response strategy. We, however, found the response biases to be very similar,  $\beta_{L1} = 0.90$ ,  $\beta_{L2} =$

0.87; and  $c_{L1} = -0.10$ ,  $c_{L2} = -0.15$ ,  $\text{Cohen } d = -0.64 [-1.80, 0.52]^3$ . Lower  $\beta$ -value suggests a more conservative response criterion in L2. Negative value of the  $c$  parameter suggests participants adopted a conservative response criterion generally, but slightly more so in the L2 condition. However, the Confidence Interval of the difference between the  $c$ -parameters includes zero, hence it is statistically non-significant. Hence, changes in accuracy caused by reading news headlines in a foreign language were not convincingly accompanied by more conservative response bias.

## **Exploratory analyses**

### *Sharing*

We asked participants about their willingness to share each of the news headline they read, using 0-100 point scale. This analysis is somewhat problematic, since we asked participants to decide on sharing after they rated the believability of a news headline, which itself is an experimental manipulation (Pennycook, Epstein, et al., 2021; Pennycook & Rand, 2021b). We do not know whether any differences in declared willingness to share are a main effect of language used, or simply moderates the debiasing effects of judging the believability of a given news. Therefore, we only briefly report the results.

We submitted this score to identical LME, as we did with the believability ratings (Table 2). The results showed that people were equally willing to share false news as true news, but more so when the news was presented to them in L2. More reflective participants were less willing to share any type of news. The language effect by CRT interaction was close to significance,  $p = .052$ , indicating that cognitive reflection (somewhat) decreased the willingness to share any type of news more when headlines were presented in L1.

By the suggestion of a reviewer, we also tested whether believability ratings predict willingness to share (see Supplementary Materials). To this end, we built an LME, where we additionally controlled for type of news, language the news headline was presented, their content (News ID), and participants CRT scores. Indeed, people were more willing to share

---

<sup>3</sup> The ROCToolbox can estimate the overall response criterion  $c$  without a standard error attached, but also  $c_{1-5}$  parameters estimated for each confidence level. These parameters are estimated with standard error metric attached. We decided to compare the aggregated  $c$  parameters, using the largest SE attached to a  $c_{1-5}$  parameter (Białek et al., 2020).

news they found believable, with 1-point increase in believability increased greater willingness to share by about 4%. Critically to our interests, believability was a stronger driving force of sharing intent when news was presented in L2, compared to L1.

## **General Discussion**

Across two experiments with 570 participants - one in-lab, one online - we found that using foreign language could disrupt the ability to discern true from fake news. When using a foreign versus native language, true news was judged as equally (Study 2) or less believable (Study 1) while false news was judged as more believable (Studies 1 and 2). We observed the FLE to be robust to changes in affective processing (i.e., the effect was not mediated by perceived arousal of news, Study 1), and to reflective processing (i.e., the effect was not mediated by individual differences in CRT, Studies 1 and 2). Finally, using signal detection theory, we showed that the negative effects of using a foreign language were not caused by adopting different responding strategy (e.g., preferring omissions to false alarms in L2).

## **Practical implications**

Although many news consumers may acquire news in their foreign language<sup>4</sup>, past research has focused almost exclusively on native language consumers (and particularly in U.S. samples) (Grinberg et al., 2019; Pennycook & Rand, 2021a). Our research suggests the problem of fake news can be even larger in previously unstudied populations, which are less able to discern true from false news. If so, two groups of people become most vulnerable: (1) people interested in international news, or who seek to internationally validate their local, potentially biased news sources; and (2) immigrants. The former group may have additional difficulty in accurately contrasting their local news with international news. The latter group of people are in danger of being misinformed, decreasing their prospects in their adopted country. We therefore call scientists to devote more research to testing whether the same or different means are required to reduce the negative effects of misinformation in these populations of news consumers.

---

<sup>4</sup> English is the most commonly used language in the internet with 25.9% share (<https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/>), whereas only about 360 million people are English native speakers (about 5% out of 7.6 billion people in Earth (<https://datatopics.worldbank.org/world-development-indicators/>)).



## **Theoretical implications**

In line with recent findings showing that people tend to believe fake news more because they find it intuitive (Pennycook & Rand, 2019, 2020; Ross et al., 2019) or emotionally evocative (Martel et al., 2019), we observed that arousal and believability ratings of news go hand in hand. The effect of believing in false news was inconsistently connected with cognitive reflection scores – with more reflective individuals being better able to discern between true and false news. Finally, believability scores predicted participants intent to share news, and more so when news were presented in a foreign language.

Our results also have implications for research on the foreign language effect (FLe). Specifically, we find evidence that thinking in a foreign language is not as beneficial to the decision-maker as initially suggested (Costa et al., 2014; Keysar et al., 2012); and, in fact, that it can be detrimental. This adds to the growing literature casting doubt to the FLe as a debiasing tool (Bialek et al., 2020, 2021; Caldwell-Harris & Ayçiçeği-Dinn, 2021; Mækelæ & Pfuhl, 2019; Muda, Walker, et al., 2020; Vives et al., 2018; Xu & Liang, 2021).

We also provide some novel insight for the mechanisms that explain the FLe. Two initially proposed models suggested that foreign language use a) decreases intuitiveness via weaker affective processing or b) increases deliberation via stronger processing difficulty (Hayakawa et al., 2016). We tested both accounts, showing the detrimental effects of using foreign language was independent from the perceived arousal of news and from individual differences in cognitive reflection. Our results are consistent with studies who showed that the FLe is robust to emotions (Chan Yuen-Lai et al., 2016; Geipel et al., 2015b; Muda, Walker, et al., 2020), and does not affect cognitive reflection scores (Bialek et al., 2019, 2021; Costa et al., 2014). Hence, both the decreased-affect or increased-reflection models are not supported by our data. Our results fit to recently formulated accounts of foreign language affecting metacognition (Bialek et al., 2020; Niszczoła et al., 2022), claiming that people fail to use metacognitive cues to engage in reflection. That is, they overweight intuitively appealing responses and fail to reflect in cases when there is no salient intuitive cue. In much of the research on the psychology of judgment and decision-making, intuitions are salient and misleading. In such cases, faulty metacognition can be beneficial, because it may help people to override the intuitively appealing incorrect response. However, in everyday contexts where intuitions are often helpful, using a foreign language can be detrimental.

## **Limitation and future directions**

Best practice in language research is to use counterbalanced bilinguals, e.g., ask Spanish native speakers who know English to answer questions in Spanish (L1) or English (L2), and then English native speakers who know Spanish to answer questions in English (L1) or Spanish (L2). The observed effect of using foreign language should be visible in all L2 users, which would deconfound the effect from semantic issues in the translation or cultural context. In this research we only recruited Polish-English bilinguals, and thus we do not know whether the reported negative effect of using L2 is unique to this language pair. Moreover, we cannot determine whether our effects are attributable to language processing per se, or to foreign-language context the news were placed (Costa, Foucart, Arnon, et al., 2014 for a discussion of this issue).

The practical difficulty of such design is that English native speakers only rarely know foreign language well enough to be included in such experiments, and if they do, they are oftentimes first- or second-generation immigrants with live contact with both languages (Portes & Schauffler, 1994). Next, most bilinguals learned English as their second language, and therefore their native language must be paired with English, circling back to the lack of corresponding sample in English native speakers. Another way of avoiding such confound is to use several pairs of languages always paired with English as done in Białek, Domurat, et al. (2021); Białek et al. (2019); Costa, Foucart, Arnon, et al. (2014); Hayakawa et al. (2019). Future research could find such language pairs of languages and its users.

Finally, for methodological reasons we only sampled bilingual participants who are not nor were for extensive period of time living abroad. Hence, we believe extending our findings on such a group is problematic, but not necessarily wrong.

## **Conclusions**

We show that bilinguals using their second language are less able to discern true from fake news. This effect is driven by both decreased believability of true and increased believability of false news. We also show that this effect is not moderated by language effect on perceived arousal of news nor by changes in cognitive reflection. These results highlight that people using their second language, potentially including refugees and economic immigrants who are especially vulnerable groups, can be additionally harmed by fake news. If

further studies (preferably field studies on social media, Mosleh et al., 2021) confirm this phenomenon, methods mitigating this effect should be urgently developed.

## **Acknowledgements**

The current project was financed by the resources of Polish National Science Centre (NCN) assigned by the decision no. PRELUDIUM 2018/29/N/ HS6/02058 to RM. Work done by MB was supported by the National Science Centre, Poland (NCN) under Grant no. SONATA 2017/26/D/HS6/ 01159. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## **Authorship contributions**

Rafał Muda: Conceptualization, Methodology, Investigation, Validation, Statistical analysis, Writing - original draft, Writing - review & editing, Funding acquisition, Project administration

Gordon Pennycook: Conceptualization, Methodology, Writing - review & editing

Damian Hamerski: Methodology, Investigation, Writing - review & editing

Michał Białek: Conceptualization, Methodology, Validation, Statistical analysis, Writing - original draft, Writing - review & editing, Funding acquisition, Supervision

## References

- Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A Signal Detection Approach to Understanding the Identification of Fake News. *Perspectives on Psychological Science*, 17(1), 78–98. <https://doi.org/10.1177/1745691620986135>
- Białek, M., Domurat, A., Paruzel-Czachura, M., & Muda, R. (2021). Limits of the foreign language effect: Intertemporal choice. *PsyArXiv*. <https://doi.org/10.31219/osf.io/vhp86>
- Białek, M., Domurat, A., Paruzel-Czachura, M., & Muda, R. (2022). Limits of the foreign language effect: Intertemporal choice. *Thinking & Reasoning*, 28(1), 97–124. <https://doi.org/10.1080/13546783.2021.1934899>
- Białek, M., Muda, R., Stewart, K., Niszczoła, P., & Pieńkosz, D. (2020). Thinking in a foreign language distorts allocation of cognitive effort: Evidence from reasoning. *Cognition*, 205, 104420. <https://doi.org/10.1016/j.cognition.2020.104420>
- Białek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, 85, 103855. <https://doi.org/10.1016/j.jesp.2019.103855>
- Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2021). When using the native language leads to more ethical choices: Integrating ratings and electrodermal monitoring. *Language, Cognition and Neuroscience*, 36(7), 885–901. <https://doi.org/10.1080/23273798.2020.1818266>
- Chan Yuen-Lai, Xuan, G., Ng, J. C., & Tse Chi-Shing. (2016). Effects of dilemma type, language, and emotion arousal on utilitarian vs deontological choice to moral

- dilemmas in Chinese-English bilinguals. *Asian Journal of Social Psychology*, 19(1), 55–65. <https://doi.org/10.1111/ajsp.12123>
- Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). “Piensa” twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254.
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28(5), 503–509.
- Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It’s a response bias effect. *Psychological Review*, 117(3), 831–863.  
<https://doi.org/10.1037/a0019634>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241.  
<https://doi.org/10.1177/1745691612460685>
- Fernández-López, M., & Perea, M. (2020). Language does not modulate fake news credibility, but emotion does. *Psicológica Journal*, 41(2), 84–102.  
<https://doi.org/10.2478/psicolj-2020-0005>
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, 59, 8–17.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). The foreign language effect on moral judgment: The role of emotions and norms. *PloS One*, 10(7), e0131529.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378. <https://doi.org/10.1126/science.aau2706>

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a Foreign Language Changes Our Choices. *Trends in Cognitive Sciences*, 20(11), 791–793.

Hayakawa, S., Lau, B. K. Y., Holtzmann, S., Costa, A., & Keysar, B. (2019). On the reliability of the foreign language effect on risk-taking. *Quarterly Journal of Experimental Psychology*, 72(1), 29–40. <https://doi.org/10.1177/1747021817742242>

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668.

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2017). The ROC Toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior Research Methods*, 49(4), 1399–1406. <https://doi.org/10.3758/s13428-016-0796-z>

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

- Lee, T. (2019). The global rise of “fake news” and the threat to democratic elections in the USA. *Public Administration and Policy*, 22(1), 15–24. <https://doi.org/10.1108/PAP-04-2019-0008>
- Mækelæ, M., & Pfuhl, G. (2019). Deliberate reasoning is not affected by language. *PLOS ONE*, 14(1), e0211428. <https://doi.org/10.1371/journal.pone.0211428>
- Martel, C., Pennycook, G., & Rand, D. G. (2019). *Reliance on emotion promotes belief in fake news* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/a2ydw>
- Montero-Melis, G., Isaksson, P., van Paridon, J., & Ostarek, M. (2020). Does using a foreign language reduce mental imagery? *Cognition*, 196, 104134. <https://doi.org/10.1016/j.cognition.2019.104134>
- Mosleh, M., Pennycook, G., & Rand, D. G. (2021). Field experiments on social media. *Current Directions in Psychological Science*, 09637214211054761.
- Muda, R., Pieńkosz, D., Francis, K., & Białek, M. (2020). The moral foreign language effect is stable across presentation modalities. *Quarterly Journal of Experimental Psychology*, 174702182093507. <https://doi.org/10.1177/1747021820935072>
- Muda, R., Walker, A. C., Pieńkosz, D., Fugelsang, J. A., & Białek, M. (2020). Foreign Language does not Affect Gambling-Related Judgments. *Journal of Gambling Studies*, 36(2), 633–652. <https://doi.org/10.1007/s10899-020-09933-6>
- Niszczoła, P., Pawlak, M., & Białek, M. (2022). Bilinguals are less susceptible to the bias blind spot in their second language. *PsyArXiv*. <https://doi.org/10.31234/osf.io/9vu76>



- Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A Practical Guide to Doing Behavioral Research on Fake News and Misinformation. *Collabra: Psychology*, 7(1), 25293. <https://doi.org/10.1525/collabra.25293>
- 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(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015a). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24(6), 425–432.
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015b). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, 80, 34–72.
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 88(2), 185–200. <https://doi.org/10.1111/jopy.12476>
- Pennycook, G., & Rand, D. G. (2021a). The psychology of fake news. *Trends in Cognitive Sciences*, *in press*.
- Pennycook, G., & Rand, D. G. (2021b). *Reducing the spread of fake news by shifting attention to accuracy: Meta-analytic evidence of replicability and generalizability* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/v8ruj>

Portes, A., & Schauffler, R. (1994). Language and the Second Generation: Bilingualism Yesterday and Today. *International Migration Review*, 28(4), 640–661.  
<https://doi.org/10.1177/019791839402800402>

Romero-Rivas, C., Corey, J. D., Garcia, X., Thierry, G., Martin, C. D., & Costa, A. (2017). World knowledge and novel information integration during L2 speech comprehension\*. *Bilingualism: Language and Cognition*, 20(3), 576–587.  
<https://doi.org/10.1017/S1366728915000905>

Ross, R. M., Rand, D. G., & Pennycook, G. (2019). *Beyond “fake news”: The role of analytic thinking in the detection of inaccuracy and partisan bias in news headlines* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/cgsx6>

The Jamovi Project. (2021). *Jamovi 2.2.5*. <https://www.jamovi.org/>

Trippas, D., Verde, M. F., & Handley, S. J. (2014). Using forced choice to test belief bias in syllogistic reasoning. *Cognition*, 133(3), 586–600.  
<https://doi.org/10.1016/j.cognition.2014.08.009>

Vives, M.-L., Aparici, M., & Costa, A. (2018). The limits of the foreign language effect on decision-making: The case of the outcome bias and the representativeness heuristic. *PLOS ONE*, 13(9), e0203528. <https://doi.org/10.1371/journal.pone.0203528>

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(2), 201.

Xu, Y., & Liang, Z.-Y. (2021). *The foreign-language discount effect: Using English increases intertemporal discount rates through more distant future perception* [Preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/vrtf4>

