

Proficient bilinguals make less superstitious choices revisiting the FLE in light of L2 proficiency

Silvia Purpuri¹ , Nicola Vasta¹ , Barbara Treccani¹, Roberto Filippi² and Claudio Mulatti¹

¹University of Trento, Italy and ²University College London IOE Faculty of Education and Society, UK

Research Article

Cite this article: Purpuri, S., Vasta, N., Treccani, B., Filippi, R. and Mulatti, C. (2026). Proficient bilinguals make less superstitious choices revisiting the FLE in light of L2 proficiency. *Bilingualism: Language and Cognition* 1–9. <https://doi.org/10.1017/S1366728926101643>

Received: 13 September 2025 UTC

Revised: 23 June 2026 UTC

Accepted: 24 June 2026 UTC

Keywords:

bilingualism; L2 proficiency; Foreign Language Effect; FLE; superstitious choice

Corresponding author:

Silvia Purpuri;

Email: silvia.purpuri@unitn.it

Abstract

The relationship between language and other domains of cognition is a key area of interest in bilingualism research. A relevant phenomenon in this regard is the Foreign Language Effect (FLE), which suggests that individuals may make more rational decisions when using a second language (L2) compared to their native language (L1). Here, by replicating a previous study, we investigated how different levels of L2 proficiency influence superstitious beliefs among bilingual individuals, beyond the language (L1 or L2) in which they are tested. Our findings showed no significant differences in superstition levels between the L1 and L2 contexts. However, we found that greater L2 proficiency was linked to lower superstition levels in both languages. These results suggest that L2 proficiency, or variables associated with it, plays a key role in shaping rational decision-making and may lead to inaccurate estimates of the FLE if it is not adequately controlled for. This study advances our understanding of the FLE by emphasising the importance of considering L2 proficiency in research on bilingual cognition and judgement.

Highlights

- No significant Foreign Language Effect emerged on superstition beliefs.
- L2 proficiency, not test language, predicted lower superstition levels.
- Higher L2 proficiency was linked to more rational judgments.
- L2 proficiency should be controlled in Foreign Language Effect research.

1. Introduction

In an increasingly multilingual world, understanding how language influences cognition is crucial. Research on the Foreign Language Effect (FLE) comparing bilinguals using their native (L1) and foreign (L2) language suggests that using a foreign language can influence moral judgement and decision-making (for a review, see Circi et al., 2021; Purpuri et al., 2025b). In a typical FLE experiment, the experimental group performs a task in L2 that taps into moral and/or decision-making processes, while the control group performs an equivalent version of the task in L1. Results generally reveal systematic differences between the responses of the two groups. While much attention has been paid to these language-dependent shifts in higher-order evaluative processes, one crucial factor has often been overlooked: how proficient someone is in L2.

Differences in L2 proficiency between the experimental and control groups could constitute a confounding factor in this type of studies. In general, random assignment allows researchers to achieve a (probabilistic) balance of pre-existing participant characteristics (e.g., L2 proficiency, I.Q., etc.) across conditions. However, random assignment alone is effective only when sample sizes are sufficiently large, a condition that psychological experiments rarely meet (cf. Goldberg, 2019). An imperfect and uncontrolled balancing of confounding variables, when coupled with publication bias (i.e., the pervasive tendency in the psychological literature for null or opposite-to-expected findings not to be published (cf. de Bruin et al., 2015a, De Bruin et al., 2015b), may lead to a distorted representation of the literature on a given phenomenon. This risk is also present in the literature on the FLE (cf. Hu et al., 2026), raising the concrete possibility that this body of research may not faithfully reflect the true effect of using L2 on reasoning processes.

In addition to imperfectly effective randomisation, systematic factors may also contribute to a differential representation of L2 proficiency levels across experimental and control groups. A particularly relevant factor in this respect is differential dropout related to language proficiency. Online FLE studies, including previous work from our lab (e.g., Nadarevic et al., 2018; Purpuri et al., 2024b, Purpuri et al., 2025c), have reported higher dropout rates in the experimental (L2-tested) condition than in the control (L1-tested) condition. Moreover, in some studies (e.g., Hadjichristidis et al., 2019), more participants have been assigned to the experimental condition

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

to compensate for anticipated data loss due to miscomprehension of the experimental materials. This pattern suggests that participants with insufficient L2 proficiency may be more likely to self-exclude when experimental materials are presented in L2 (i.e., in the experimental condition). Consequently, despite random assignment, the experimental group may disproportionately comprise individuals with higher L2 proficiency compared to the control group.

Actually, the potential moderating role of L2 proficiency in the FLE is not a new concern, and some studies have attempted to address it often reporting mixed results, with some studies finding moderating effects of L2 proficiency and others not (for a review, see Del Maschio *et al.*, 2022; Pavlenko, 2017; Purpuri *et al.*, 2025b). In general, however, studies addressing this issue have mostly relied on subjective measures of proficiency (cf. Del Maschio *et al.*, 2022), assessed L2 proficiency only in the experimental group and/or failed to measure proficiency comprehensively, that is, through tests covering multiple linguistic domains. Moreover, even when L2 proficiency was assessed, it was often not directly included in the analyses, but rather used only to verify matching between the experimental and control groups on this variable.

For instance, Privitera *et al.* (2023) assessed L2 proficiency in both the experimental and control groups using an objective test (i.e., 15 grammar questions and 10 vocabulary questions). However, despite showing that the control and experimental groups had comparable levels of L2 proficiency, they ultimately chose not to include these measures in their analyses. Moreover, the inclusion of objective L2 proficiency measures led to the exclusion of eight participants from the experimental group due to voluntary drop-outs, which may again point to the self-selection bias described above. Similarly, Del Maschio *et al.* (2022) assessed L2 proficiency using a 40-item questionnaire (30 grammar and 10 comprehension questions), but only for the experimental group and not for the group assigned to the native language condition. The current study addresses the fact that previous research on the FLE has not typically collected objective measures of L2 proficiency and, when such measures were collected, did not incorporate them into the analysis, or administered them to both experimental and control groups.

This limitation is particularly relevant given that L2 proficiency may influence the extent to which individuals process information in a foreign language. Studies on the FLE show that using and thinking in a foreign language can dampen emotional intensity, pushing people to make more utilitarian decisions. For example, in foreign language, people might take more risks or be more willing to accept harm if it maximises overall outcomes (Keysar *et al.*, 2012; Winskel and Bhatt, 2020). It can also reduce superstitious thinking (Hadjichristidis *et al.*, 2019) and change how we perceive dishonesty and crime (Alempaki *et al.*, 2021). Recent evidence also shows that processing information in a foreign language increases tolerance of ambiguity, the degree to which people can tolerate novel and/or unpredictable situations, compared to a native language (Purpuri *et al.*, 2024b). Yet, how much does proficiency in that foreign language really matter in these scenarios?

Language proficiency is more than just speaking a language; it is about mastering a complex web of skills, from everyday conversation to the cognitive demands of academic tasks (Purpuri *et al.*, 2024a; Purpuri *et al.*, 2025a). Frameworks like the Common European Framework of Reference for Languages (CEFR) break it down into levels (A1 to C2), but scholars debate whether such classifications fully capture the depth of proficiency, especially when it comes to higher-order thinking (Hulstijn, 2011).

When we consider how language affects decision-making, proficiency becomes a key piece of the puzzle. A lack of proficiency might lead to misunderstandings or hesitation when interpreting ambiguous information, potentially skewing results. Recent research (Purpuri *et al.*, 2025c) suggests that the FLE might actually be influenced by participants' language skills, not just the switch from one language to another. In other words, differences in fluency could be the real driver behind these cognitive shifts, rather than the language itself.

1.1. The present study

The present study aims to further investigate the connection between L2 proficiency and superstition, building on our previous work (Purpuri *et al.*, 2024b, Purpuri *et al.*, 2025c) and drawing inspiration from the material and experiment on superstition conducted by Hadjichristidis *et al.* (2019).

Hadjichristidis *et al.* (2019) showed that thinking in a foreign language might reduce superstitious beliefs, suggesting that using a non-native language could attenuate intuitive or heuristic thinking. However, their study relied solely on self-reported measures of L2 proficiency and, importantly, assessed L2 proficiency only in the experimental group and not in the control group, leaving open the question of whether the effect was driven by the novelty of the language context or by individuals' ability to process it fluently.

By including objective measures of L2 proficiency, we were aiming for a clearer understanding of how language and fluency intersect with superstitious thinking. We hypothesised that higher L2 proficiency will lead to less superstitious thinking, no matter the language of the test. This has important implications for everything from education to cognitive therapy, as we continue to unravel the intricate ways language relates to cognition.

To test our hypothesis, we conducted a large online study involving responses from a total of 355 participants. All participants were asked to evaluate four superstitious scenarios and four filler scenarios, indicating whether the meaning of the scenario was positive or negative. Half of the participants read these scenarios in their native language (Italian), while the other half read them in a foreign language (English). Additionally, all participants completed an extended English language placement test and answered questions about their magical beliefs (see Methods). This approach allowed us to disentangle the individual contributions of the FLE and proficiency in the foreign language on participants' superstitious beliefs.

While the present study is not the first to address the issue of L2 proficiency in the FLE, it seeks to contribute to the literature by adopting a more comprehensive and fine-grained assessment of second language proficiency. Specifically, L2 proficiency was measured using a 100-item test covering levels from A1 to C2, allowing for a more detailed characterisation of participants' linguistic competence across a broad proficiency spectrum.

This approach differs from much of the existing literature, where proficiency has often been assessed with relatively limited objective measures (e.g., short grammar and vocabulary tests) or through subjective self-report scales. For instance, prior studies have employed brief instruments consisting of a small number of items, which may not fully capture the multidimensional nature of language proficiency (e.g., Privitera *et al.*, 2023). Moreover, as highlighted in previous meta-analytic work (e.g., Del Maschio *et al.*, 2022), the measurement of L2 proficiency has frequently been inconsistent across studies and, in some cases, not clearly reported for all experimental groups.

The present study therefore prioritises a more robust operationalisation of L2 proficiency to increase the precision with which its role in modulating the FLE can be examined. At the same time, we acknowledge that other dimensions of language experience (e.g., frequency of use, age of acquisition, immersion) may also influence the effect. These factors were not directly assessed in the current design and should be systematically considered in future research to provide a more holistic account of bilingual experience.

2. Methods

2.1. Participants

The sample size was determined using G*power. Based on Study 3 of Hadjichristidis et al., 2019, we performed a power analysis using the effect size of the Language $\times$ Scenario type interaction (i.e., $f = .16$). In order to obtain a power = 0.80 with an alpha = .05, the analysis indicated a minimum sample size of $n = 309$ participants for an analysis of covariance (ANCOVA; two groups, one covariate), with half of the participants being assigned to one condition (L1) and the other half to the second condition (L2) (see pre-registration: <https://osf.io/y85vd>). Nevertheless, to maximise the study's statistical power and account for potential dropouts, we opted to gather data from as many participants as feasible.

A total of 431 volunteers participated in the study and completed the experiment online. They also provided biographical and linguistic information. After eliminating incomplete answers and two participants who completed the questionnaire in less than 5 minutes (an insufficient amount of time to meaningfully complete the questionnaire), 355 participants were included in the analysis (249 female, 100 male, 4 non-binary, 2 did not provide this information); mean age = 21.66 years (1 participant did not provide this information, age range: 18–56 years). Then, 156 participants were randomly assigned to the Italian questionnaire (L1_Italian condition), and 199 to the questionnaire in English (L2_English condition). All participants had studied English as a second language as part of their previous scholastic education. Participants were asked to self-rate their level of English according to the Common European Framework. They could choose between B1, B2, C1 and C2. Then, 111 participants chose B1, 146 participants B2, 46 participants C1 and 25 participants C2, while 27 did not provide this response. Overall, 131 participants stated that they possess an official language certificate: 21 reported that they possess a B1 certification, 65 a B2 certification, 30 a C1 certification and 15 a C2 certification.

All participants completed an English language placement test. Five participants positioned themselves at an elementary/pre-intermediate (A1/A2) level, 62 participants at an intermediate (B1) level, 158 participants at an upper intermediate (B2) level, 125 participants at an advanced (C1) level, and five participants at a proficient (C2) level (see Table 2).

Participants also reported their level of education, with 32 participants holding a middle school diploma, 294 a secondary school diploma, nine a bachelor's degree, 16 a master's degree, five a doctoral degree and three a master's degree.

The study (Ref. 2024-04 ESA) received ethical approval from the Ethics Committee of the University of Trento and participants provided their consent at the beginning of the online questionnaire. Informed consent was obtained from all participants. Research has been performed in accordance with the Declaration of Helsinki (<https://www.wma.net/policies-post/wma-declaration-of-helsinki/>).

2.2. Materials and procedure

At the initial stage, we used the same materials and followed the same procedure as in the original experiment (Hadjichristidis et al., 2019, Study 3). The study was conducted by means of an online survey administered through Qualtrics. Participants evaluated eight scenarios: four superstitious scenarios (two bad-luck and two good-luck scenarios) and four filler scenarios (see Table 1). The scenarios were presented in two blocks, each comprising two superstitious scenarios (either bad luck or good luck) and two filler scenarios. For the block containing the bad-luck scenarios (and two filler scenarios), we used a 9-point scale ranging from 1 (neutral) to 9 (very negative), whereas for the block containing the good-luck

Table 1. English version of the materials used in the experiment

Scenarios	Description
Bad-luck scenarios	
Mirror	Imagine that you have surgery at the hospital. On the day of the surgery, your bathroom mirror breaks. How would you feel about having the surgery under this condition?
Black cat	Imagine that you have to take your practical driving test, which you expect to be difficult. Just before your driving test, a black cat crosses your path from left to right. How would you feel about taking your driving test under this condition?
Good-luck scenarios	
Four-leaf clover	Imagine that you have an important presentation at work and you feel nervous. If the presentation goes well, it could lead to a promotion. As you practise your talk at a park nearby, you find a four-leaf clover in the grass. How would you feel about giving the presentation under this condition?
Falling star	Imagine that you are going to a party, and someone you like will also be there. You are not sure yet whether this person has feelings for you. It is late in the evening, and you have just arrived at the party. You look up at the sky and see a falling star. How would you feel meeting the person you like under this condition?
Control scenarios	
Blocked sink	Imagine that you have to submit a job application. On the day of the submission, you notice that the kitchen sink is blocked. How would you feel about submitting the application under this condition?
White dog	Imagine that you have to take an exam. Before you take the exam, you see a man with a white dog standing in front of the building where the exam will take place. How would you feel about taking the exam under this condition?
Tulips	Imagine that you are searching for a new apartment to rent. You found a nice one and on the way to hand in the documents which the owner requested. While walking towards the rental agency, you notice a public garden with tulips. How would you feel about handing in your forms under this condition?
Airplane	Imagine that you have a job interview, and you are well prepared. Just before the interview, you have a coffee. While walking back from the café to the location where the interview will take place, you look up to the sky and see an airplane. How would you feel about giving the job interview under this condition?

Table 2. Common European Framework of Reference for Languages (CEFR)

CEFR level			Objective testing (number of participants per level)	Self-rated (number of participants per level)
Beginner	A1	Elementary	5	0
	A2	Pre-intermediate		
Intermediate	B1	Intermediate	62	111
	B2	Upper intermediate	158	146
Advanced	C1	Advanced	125	46
	C2	Proficient	5	25

Note: This guideline serves to describe the achievements of learners of foreign languages across Europe and, increasingly, in other countries. Developed by the Council of Europe as the central component of the project “Language Learning for European Citizenship” between 1989 and 1996, its primary goal is to offer a universal method for learning, teaching and assessing languages across Europe. The six reference levels are gaining widespread recognition as the European standard for assessing an individual’s language proficiency.

scenarios (and two filler scenarios), we used a 9-point scale ranging from 1 (neutral) to 9 (very positive). As in the original experiment, we supplemented the scales with SAM figurines (Lang, 1985). The presentation order of blocks and the scenarios was randomised separately for each participant. The scenarios were presented one at a time, each on a separate page. Once participants judged a scenario and moved to the next one, they could not revert to previous pages to make corrections.

After completing the test, adding to the original experiment, we administered a complete English language placement test, prepared by an English as a Foreign Language teacher, with adaptations from official certification tests. The 100 questions of the tests ranged from an A1 to a C2 level (see Table 2 for a description of the levels and the number of participants for each level) and aimed to evaluate participants’ L2 proficiency.

Participants were subsequently asked to complete the magical beliefs scale used in the original experiment (adapted from Wisemann and Watt, 2004), which they all received in their native language. The magical beliefs scale included six items: “Do you say “fingers crossed” or actually cross your fingers?” “Do you sometimes carry a lucky charm or object?” “Do you sometimes knock on wood or touch wood?” “Would you be anxious about breaking a mirror because it is thought to cause bad luck?” “Are you superstitious about the number 13?” “Have you avoided walking under a ladder because it is associated with bad luck?” Participants provided responses to each question using a 5-point scale, ranging from 1 (No, absolutely not) to 5 (Yes, absolutely). The inclusion of this scale aimed to investigate whether language influences feelings evoked by superstitious scenarios beyond individual differences in magical beliefs.

Finally, participants filled out a set of demographic questions, where they were asked to provide information about their cultural and linguistic background.

3. Results

3.1. Effect of language and L2 proficiency on superstition

An ANCOVA was carried out to test if there was an effect of language and L2 proficiency on the questionnaire’s answers. We included the superstitious scenarios’ scores (Questions 1 – mirror,

2 – black cat, 5 – four-leaf clover and 6 – falling star) as the dependent variable, language as the fixed factor and L2 proficiency as the covariate. The main effect of language did not reach the significance criterion ($F = 0.14$), while L2 proficiency resulted significant, $F(1, 352) = 4.68$, $p = .031$, $\eta^2 = .013$, revealing that, regardless of the language in which the questionnaire was presented, the more proficient participants were in the second language, the lower the value assigned to the items, and thus the lower their level of superstition (Figure 1).

We conducted the same ANCOVA on filler scenario scores, with language as the fixed factor and L2 proficiency as the covariate. In this analysis, neither the main effect of language nor L2 proficiency was significant ($F = 0.30$ and $F = 0.68$, respectively), suggesting that L2 proficiency affected only superstitious scenarios.

3.2. Magical belief and superstition

Next, we considered the Magical Belief Scale that the participants had to fill in. We conducted an ANCOVA with the superstitious scenarios’ scores as the dependent variable, language as a fixed factor, and the scores on the Magical Belief Scale as a covariate. The covariate was found to be significant ($F(1, 325) = 154.84$, $p < .001$, $\eta^2 = .323$), indicating that superstitious participants (i.e., lower scores on the Magical Belief Scale) responded more superstitiously in the superstitious scenarios. We conducted the same ANCOVA on filler scenarios, with language as a fixed factor and the magical belief scale as a covariate. Again, scores on the Magical Belief Scale were significant ($F(1, 325) = 17.08$, $p < .001$, $\eta^2 = .050$), indicating that, although to a lesser extent than for the superstitious scenarios, more superstitious participants continued to respond more superstitiously even to filler scenarios. These results suggest that superstitious participants tend to assign emotional value to all types of scenarios, including neutral ones (Figure 2).

3.3. Relationship between L2 proficiency and magical belief

To explore if there is a direct relation between proficiency in L2 and magical belief, we conducted a Pearson’s correlation analysis between L2 proficiency and Magical Belief Scale. This resulted in a positive correlation ($r = .12$, $p = .029$), indicating that the more proficient participants are in English, the more their responses tend towards 5 in the Magical Belief Scale, which is the value indicating the least superstitious (see Figure 3). This means that as L2 proficiency increases, participants are more likely to score towards the “least superstitious” end of the scale, reinforcing the idea that language proficiency plays a role in reducing superstitious beliefs.

3.4. Impact of age on superstitious beliefs

In an ANCOVA, we examined whether Age had an effect on superstition. One participant was excluded because they did not report their age. We set the superstitious scenarios as dependent variables, language as a fixed factor and age as a covariate. Age resulted in a significant covariate, $F(1, 351) = 5.37$, $p = .021$, $\eta^2 = .015$, showing that older participants were less superstitious if compared to younger participants (Figure 4). Previous research has demonstrated a positive association between age and superstitious beliefs (Epstein, 1993). However, other studies have found a different pattern, indicating that younger individuals tend to be more superstitious than older adults (Corrigan et al., 1980; Gallup and Newport, 1991).

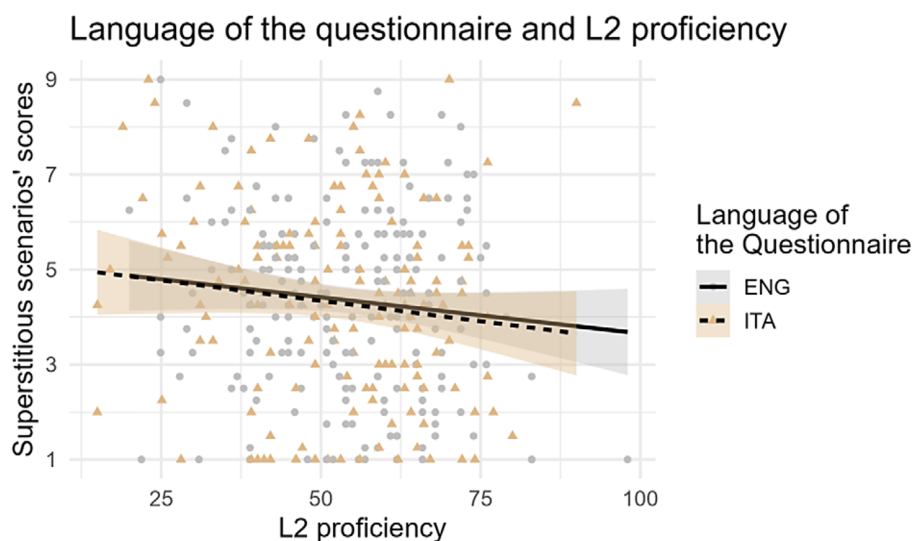


Figure 1. Superstitious scenarios' scores (mean) sorted by L2 proficiency, in both the English (ENG) and Italian (ITA) conditions. Each dot corresponds to the score of at least one participant. Grey and brown regions indicate the confidence interval (95%).

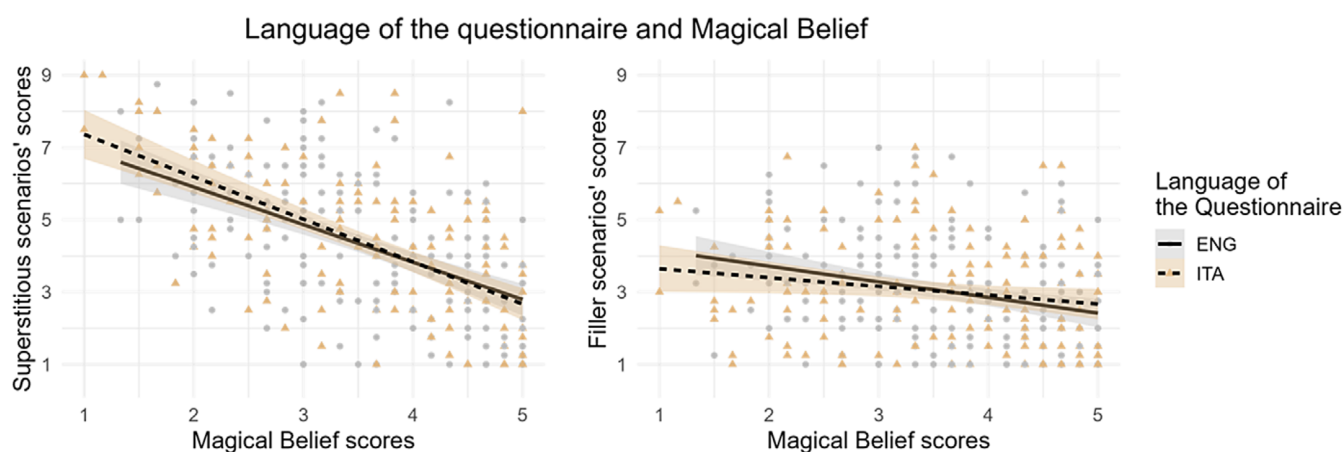


Figure 2. Superstitious and filler scenarios' scores (mean) sorted by Magical Belief scores, in both the English (ENG) and Italian (ITA) conditions. Each dot corresponds to the score of at least one participant. Grey and brown regions indicate the confidence interval (95%).

3.5. Gender differences in superstitious thinking

We then performed another ANOVA in order to examine if there were any gender-related differences on superstition. Six participants were excluded from this analysis (4 = non-binary, 1 = other, 1 participant did not provide this information), as their number was too small to be consistent. We set the superstitious scenarios as dependent variables, language and gender as fixed factors. The main effect of gender resulted to be significant, $F(1, 345) = 33.30$, $p < .001$, $\eta^2 = .088$, with female participants being more superstitious (mean = 4.705, $sd = 1.85$) if compared to male participants (mean = 3.393, $sd = 1.89$), while the interaction between Gender and Language of the questionnaire did not reach the significance criterion ($F = 0.79$; see Figure 5). While females scored an average of 4.7 on the superstitious scenarios, males scored 3.4, suggesting a noticeable difference in how superstition is experienced between genders. The absence of an interaction between gender and the language of the questionnaire indicates that these gender differences are independent of the language used in the study (Figure 5).

3.6. Educational level and superstition

Next, in an ANCOVA, we examined whether the reported level of instruction had an effect on superstition. However, contrary to expectations, in this analysis, the level of instruction did not emerge as a significant covariate ($F = 0.45$, $p = .50$), indicating that formal education does not play a major role in shaping superstitious beliefs in this context.

4. Discussion

The primary aim of the present study was to determine whether higher proficiency in L2 influences participants' superstitious beliefs, regardless of the language in which the questionnaire is administered. The study investigated the role of second language proficiency on superstitious beliefs, challenging the conventional understanding of the FLE.

Our analysis of data from 355 participants revealed no significant differences between responses given in the participants' native

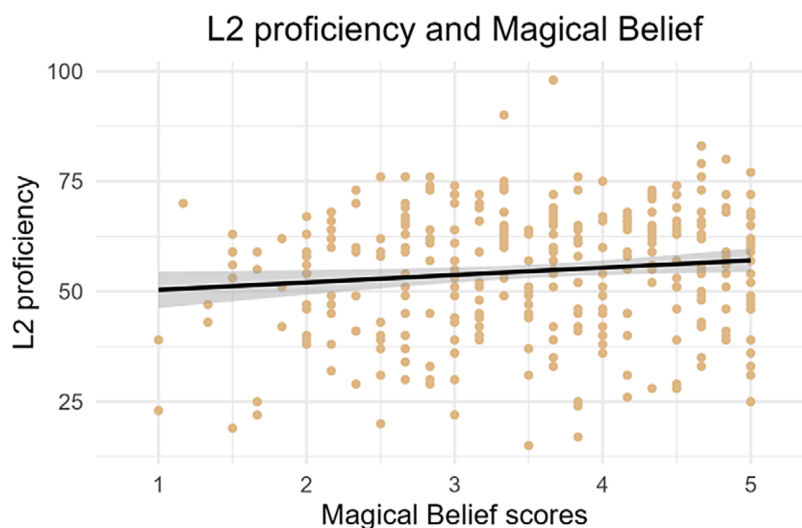


Figure 3. L2 proficiency sorted by Magical Belief scores. Each dot corresponds to the score of at least one participant. The grey region indicates the confidence interval (95%).



Figure 4. Superstitious scenarios' scores (mean) sorted by age, in both the English (ENG) and Italian (ITA) conditions. Each dot corresponds to the score of at least one participant. Grey and brown regions indicate the confidence interval (95%).

language and those given in their second language. These findings do not provide evidence in support of the FLE hypothesis, according to which using a foreign language should result in more rational and less emotionally driven responses. Instead, our results showed that L2 proficiency played a critical role: participants with higher proficiency in their second language exhibited lower levels of superstition, regardless of whether they completed the questionnaire in L1 or L2. Most previous studies' findings (e.g., Hadjichristidis *et al.*, 2019) suggest that using an L2 leads to more rational thinking due to emotional distancing. In contrast, our results indicate that L2 proficiency itself is a more significant predictor of reduced superstitious beliefs. Participants with higher English proficiency demonstrated less superstition, irrespective of the language of the questionnaire. These findings are also inconsistent with our pre-registered hypotheses, in which we expected L2 proficiency to moderate the FLE. Instead, we found an effect of L2 proficiency, but no evidence of an FLE. Whereas previous studies have almost

consistently reported the presence of the FLE across different contexts, the pattern observed here suggests that the FLE may be less robust and less pervasive than previously assumed, raising the possibility that earlier estimates of the FLE have been partly distorted by publication bias (cf. Hu *et al.*, 2026).

The lack of significant language effects on superstition observed in the present study and the prominent role of L2 proficiency also highlight the need to reconsider the results of previous FLE's studies. Previous studies may have overlooked the impact of L2 proficiency by focusing solely on the language in which materials were presented or by relying solely on subjective measures to assess proficiency (e.g., participants' self-assessment of their English proficiency). Differences in L2 proficiency between experimental and control groups, when unassessed or inadequately measured, could artificially generate an FLE, even when the observed effects are actually driven by proficiency disparities. Objective measures of language proficiency, such as standardised language tests, are

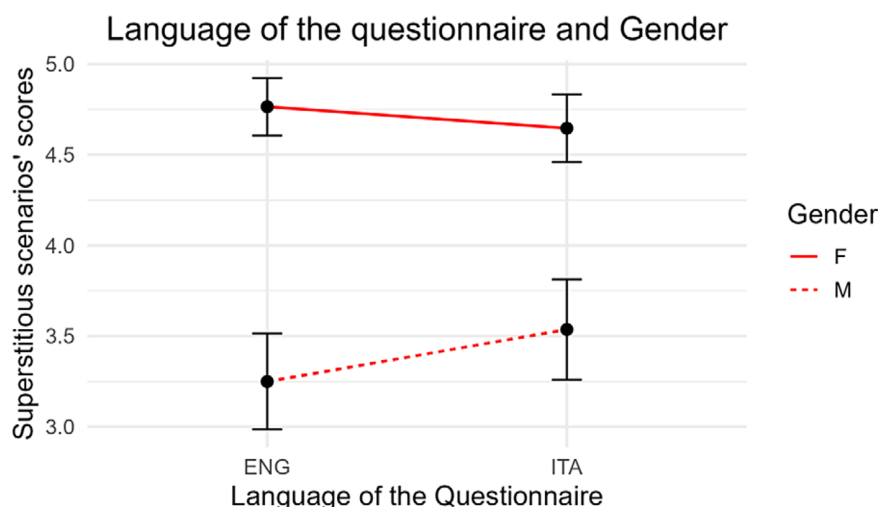


Figure 5. Superstitious scenarios' scores (mean), sorted by Language of the questionnaire and Gender (F = female, M = male). Vertical bars represent standard errors, while black dots represent the aggregated means.

essential to ensure a more accurate assessment (Privitera et al., 2023, 2025a).

In Table 2, it is possible to see a comparison between the number of participants at each English proficiency level as determined by objective CEFR testing and self-reported ratings. Notably, there are significant discrepancies between the two methods of assessment. For instance, no participants self-rated as beginners (A1), yet objective testing identified five individuals at this level. Conversely, while 62 participants objectively tested at B1 (Intermediate), 111 participants self-identified at this proficiency level. The upper-intermediate level (B2) shows more consistency, with 158 participants objectively tested and 146 self-rated in this category. A striking divergence is observed at the advanced levels. Objective testing classified 125 participants at C1 (Advanced), while only 46 self-reported at this level. Similarly, at the C2 (Proficient) level, only five participants were objectively tested, but 25 self-identified as proficient. These discrepancies highlight potential overestimation or underestimation of language proficiency by participants when self-reporting. It underscores the importance of using objective measures to accurately assess language skills, as self-assessment may not reliably reflect actual proficiency. This misalignment can have significant implications for research relying on self-reported data, which constitute the majority of studies assessing L2 proficiency, suggesting that objective testing should be prioritised for more accurate classification. This shift in methodology might yield more reliable data, ultimately refining our understanding of the FLE and its implications. In the present study, analyses including the Magical Belief Scale as a covariate revealed, as expected, a significant effect, indicating that superstitious participants were more likely to make superstitious choices, thus supporting the internal coherence of our experiment. Additionally, Pearson's correlation analysis indicated that higher L2 proficiency was associated with reduced superstitious responding (i.e., less superstitious responses), suggesting once more that L2 proficiency can mitigate superstitious tendencies, although the mechanisms underlying this association remain unclear. Interestingly, participants with stronger superstitious tendencies evaluated not only superstitious scenarios, but also filler scenarios, as more superstitious. This may reflect a general tendency among more superstitious individuals to perceive a wider range of situations as capable of influencing the

outcome of their choices, regardless of whether they are broadly recognised as superstitious. Importantly, however, the effect observed for filler scenarios was substantially smaller than that found for superstitious scenarios ($\eta p^2 = .050$ versus $\eta p^2 = .323$).

Although the FLE has often been interpreted as evidence that using a foreign language reduces emotional reactivity and leads to less biased and more rational decision-making, this account has been increasingly challenged in recent literature. In particular, the assumption that foreign language use systematically improves decision quality does not hold consistently across domains.

For example, Hayakawa et al. (2019) reported unreliable effects of foreign language use on risk-taking and raised concerns about the robustness of earlier findings. Extending this line of work, Muda et al. (2025) found that foreign language use increased risk-taking in prenatal testing decisions, a context in which higher risk has meaningful real-world consequences, and where attenuated emotional reactivity did not account for the observed pattern. Similarly, Muda et al. (2023) showed that individuals processing information in a foreign language were less accurate in distinguishing real from fake news. Further evidence comes from Bialek et al. (2020), who demonstrated that foreign language use distorts the allocation of cognitive effort, leading individuals to engage in unnecessary deliberation on simple problems while failing to sufficiently deliberate on problems that require it. Taken together, these findings challenge a straightforward interpretation of the FLE as promoting less biased and more rational decision-making.

In light of these inconsistencies, alternative theoretical accounts have been proposed. Hayakawa et al. (2017) framed the FLE in terms of whether foreign language use leads individuals to “think more” or “feel less,” with their findings suggesting that the effect is better explained by weakened intuitive responses rather than enhanced deliberative processing. Building on this perspective, Bialek (2023) proposed a metacognitive disruption account, according to which foreign language use interferes with conflict monitoring processes that allow individuals to detect when their initial responses may be incorrect. Consistent with this account, Hu et al. (2026) showed that foreign language use reduces feelings of rightness, which play a critical role in determining when individuals re-evaluate their moral judgments. Across these contributions, a common conclusion emerges: foreign language use does not

necessarily enhance reasoning, but may instead degrade the metacognitive signals that guide when increased cognitive effort is warranted.

The present findings are consistent with this broader framework. If foreign language use disrupts metacognitive calibration, then differences in L2 proficiency between experimental and control groups do not merely introduce random variability, but may instead produce systematic differences in the degree of metacognitive interference. From this perspective, proficiency acts not only as a moderator but also as a potential confounding factor that can mimic or amplify apparent language effects. Accordingly, the observed association between higher L2 proficiency and reduced superstition may reflect lower levels of metacognitive disruption during foreign language processing, rather than an increase in rational decision-making per se.

Our study found significant age and gender effects on superstition. Older participants tended to be less superstitious, and female participants were more superstitious compared to male participants. Previous research has reported both positive and negative associations between age and superstitious beliefs (e.g., Corrigan et al., 1980; Epstein, 1993; Gallup and Newport, 1991). Similarly, other studies have found that female participants tend to express more superstitious beliefs (e.g., Maqsood et al., 2018). However, these results may reflect cultural characteristics of our sample rather than actual effects of age or gender per se.

Regarding the participants' reported level of instruction, when examining whether it had an effect on superstition, we found that it does not significantly influence the participants' superstitious beliefs. Despite our initial assumptions that education level might play a role in moderating superstitious tendencies, our findings do not support them. Possibly, other factors might be more critical in shaping these beliefs.

One limitation of this study is that we did not assess participants' IQ or socio-economic status (SES). Therefore, the present study cannot determine which specific factor accounts for the observed effects. Indeed, the observed mitigation of superstitious tendencies may have been driven not by L2 proficiency per se, but by one of these two variables, or by other factors associated with proficiency. Individuals with higher IQ or SES are more likely to be both more proficient in L2 and less prone to superstitious beliefs. Future studies should consider this. In addition to this, our study has been developed in an experimental setting, and it might not have taken into account the importance of considering differences in foreign language experience across different contexts (cf. Privitera et al., 2023), potentially undermining the ecological validity of our results. Future studies should test our hypothesis in an applied field. One other limitation is that, while we excluded incomplete responses and participants who completed the questionnaire in an implausibly short time, we did not include an explicit comprehension check. At the same time, we note that our exclusion criteria were designed to mitigate low-quality data, and the overall completion times suggest that most participants engaged meaningfully with the task.

In conclusion, our findings suggest that higher proficiency in L2 is associated with superstitious beliefs, independently of the language used. There is a need to re-evaluate the existing literature on the FLE. Studies that have not accounted for L2 proficiency may have reported biased results, attributing effects to the language of presentation rather than to proficiency. Future research should prioritise assessing and accounting for L2 proficiency when investigating the FLE. By doing so, researchers can better isolate the effects of language from proficiency and provide a clearer understanding of

bilingual decision-making and judgement processes. Additionally, exploring other individual differences that correlate with language proficiency could offer deeper insights into the cognitive mechanisms underlying rationality and superstition.

The present findings do not allow us to conclude that L2 proficiency directly enhances rational decision-making. Rather, they suggest that proficiency-related variables may substantially shape how individuals process and respond to foreign-language materials.

Data availability statement. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- Alempaki, D., Doğan, G., & Yang, Y. (2021). Lying in a foreign language? *Journal of Economic Behavior & Organization*, *185*, 946–961. <https://doi.org/10.1016/j.jebo.2020.10.027>
- Bialek, M. (2023). Why should we study the foreign language effect: Debiasing through affecting metacognition? *Journal of Intelligence*, *11*(6), 103. <https://doi.org/10.3390/jintelligence11060103>
- Bialek, 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), 104420. <https://doi.org/10.1016/j.cognition.2020.104420>
- Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, *28*(4), 1131–1141. <https://doi.org/10.3758/s13423-020-01871-z>
- Corrigan, R. S., Pattison, L., & Lester, D. (1980). Superstition in police officers. *Psychological Reports*, *46*(3), 830–830. <https://doi.org/10.2466/pr0.1980.46.3.830>
- De Bruin, A., Treccani, B., & Della Sala, S. (2015a). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*(1), 99–107. <https://doi.org/10.1177/0956797614557866>
- De Bruin, A., Treccani, B., & Della Sala, S. (2015b). The connection is in the data: We should consider them all. *Psychological Science*, *26*(6), 947–949. <https://doi.org/10.1177/0956797615583443>
- Del Maschio, N., Crespi, F., Peressotti, F., Abutalebi, J., & Sulpizio, S. (2022). Decision-making depends on language: A meta-analysis of the foreign language effect. *Bilingualism: Language and Cognition*, *25*(4), 617–630. <https://doi.org/10.1017/S1366728921001012>
- Del Maschio, N., Del Mauro, G., Bellini, C., Abutalebi, J., & Sulpizio, S. (2022). Foreign to whom? Constraining the moral foreign language effect on bilinguals' language experience. *Language and Cognition*, *14*(4), 511–533. <https://doi.org/10.1017/langcog.2022.14>
- Epstein, S. (1993). Implications of cognitive-experiential self-theory for personality and developmental psychology. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 399–438). American Psychological Association. <https://doi.org/10.1037/10127-033>
- Gallup, G. H., & Newport, F. (1991). Belief in paranormal phenomena among adult Americans. *Skeptical Inquirer*, *15*(2), 137–146.
- Goldberg, M. H. (2019). How often does random assignment fail? Estimates and recommendations. *Journal of Environmental Psychology*, *66*, 101351. <https://doi.org/10.1016/j.jenvp.2019.101351>
- Hadjichristidis, C., Geipel, J., & Surian, L. (2019). Breaking magic: Foreign language suppresses superstition. *Quarterly Journal of Experimental Psychology*, *72*(1), 18–28. <https://doi.org/10.1080/17470218.2017.1371780>
- 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>
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, *28*(10), 1387–1397. <https://doi.org/10.1177/0956797617720944>

- Hu, Z., Martín-Luengo, B., & Navarrete, E. (2026). The impact of foreign language on meta-reasoning in moral decisions. *Journal of Cognition*, 9(1), 2. <https://doi.org/10.5334/joc.472>
- Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8(3), 229–249. <https://doi.org/10.1080/15434303.2011.565844>
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect. *Psychological Science*, 23(6), 661–668. <https://doi.org/10.1177/0956797611432178>
- Lang, P. J. (1985). The cognitive psychophysiology of emotion: Fear and anxiety. In A. H. Tuma, J. D. Maser, A. H. Tuma, & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 131–170). Lawrence Erlbaum.
- Maqsood, A., Jamil, F., & Khalid, R. (2018). Thinking styles and belief in superstitions: Moderating role of gender in young adults. *Pakistan Journal of Psychological Research*, 33
- Muda, R., Niszczoła, P., Hamerski, D., & Bialek, M. (2025). Using a foreign language increases risk-taking in prenatal testing decisions but not due to attenuated emotional responses. *Journal of Behavioral Decision Making*, 38(2), e70016. <https://doi.org/10.1002/bdm.70016>
- Muda, R., Pennycook, G., Hamerski, D., & Bialek, M. (2023). People are worse at detecting fake news in their foreign language. *Journal of Experimental Psychology: Applied*, 29(4), 712–724. <https://doi.org/10.1037/xap0000475>
- Nadarevic, L., Plier, S., Thielmann, I., & Darancó, S. (2018). Foreign language reduces the longevity of the repetition-based truth effect. *Acta Psychologica*, 191, 149–159. <https://doi.org/10.1016/j.actpsy.2018.08.019>
- Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current Opinion in Psychology*, 17, 74–78. <https://doi.org/10.1016/j.copsyc.2017.06.005>
- Privitera, A. J., Li, S., Zhou, Y., & Wang, M. (2023). Modulatory role of foreign language experience on the moral foreign language effect. *Bilingualism: Language and Cognition*, 26(5), 1038–1050. <https://doi.org/10.1017/S1366728923000275>
- Purpuri, S., Mulatti, C., Filippi, R., & Treccani, B. (2024a). Inside the kaleidoscope: Unravelling the “feeling different” experience of bicultural bilinguals. *Frontiers in Psychology*, 15, 1376076. <https://doi.org/10.3389/fpsyg.2024.1376076>
- Purpuri, S., Vasta, N., Filippi, R., Wei, L., & Mulatti, C. (2024b). The foreign language effect on tolerance of ambiguity. *Bilingualism: Language and Cognition*, 27(1), 16–24. <https://doi.org/10.1017/S1366728923000469>
- Purpuri, S., Treccani, B., Filippi, R., & Mulatti, C. (2025a). Bicultural bilinguals: Juggling languages, shifting ‘personalities’, destroying barriers. *International Journal of Bilingualism*, 29(5), 1408–1421. <https://doi.org/10.1177/13670069241292496>
- Purpuri, S., Vasta, N., Filippi, R., Treccani, B., Wei, L., & Mulatti, C. (2025b). Beyond the foreign language effect: Unravelling the impact of L2 proficiency on rationality. *Bilingualism: Language and Cognition*, 28(3), 740–747. <https://doi.org/10.1017/S1366728924000750>
- Purpuri, S., Vasta, N., Filippi, R., Wei, L., & Mulatti, C. (2025c). Does language shape the way we think? A review of the foreign language effect across domains. *International Journal of Bilingualism*, 29(1), 269–291. <https://doi.org/10.1177/13670069231225374>
- Winkel, H., & Bhatt, D. (2020). The role of culture and language in moral decision-making. *Culture and Brain*, 8(2), 207–225. <https://doi.org/10.1007/s40167-019-00085-y>
- Wisemann, R., & Watt, C. (2004). Measuring superstitious belief: Why lucky charms matter. *Personality and Individual Differences*, 37(8), 1533–1541. <https://doi.org/10.1016/j.paid.2004.02.009>