







Beyond the foreign language effect: unravelling the impact of L2 proficiency on rationality

Silvia Purpuri¹ , Nicola Vasta^{1,3} , Roberto Filippi² , Barbara Treccani¹ ,
Li Wei²  and Claudio Mulatti¹ 

¹Department of Psychology and Cognitive Science, University of Trento, Trento, Italy; ²Institute of Education, University College London, London, UK and ³Human Systems Lab, Department of Kinesiology, University of Windsor, Windsor, ON, Canada

Research Article

Cite this article: Purpuri, S., Vasta, N., Filippi, R., Treccani, B., Wei, L. and Mulatti, C. (2025). Beyond the foreign language effect: unravelling the impact of L2 proficiency on rationality. *Bilingualism: Language and Cognition* 28, 740–747. <https://doi.org/10.1017/S1366728924000750>

Received: 24 November 2023
Revised: 1 September 2024
Accepted: 11 September 2024
First published online: 6 December 2024

Keywords:

foreign language effect; FLE; core knowledge confusion; CKC; superstition; conspiracy; bilingualism; L2 proficiency; second language; foreign language

Corresponding author:

Silvia Purpuri;
Email: silvia.purpuri@unitn.it

Abstract

This study investigated the impact of reading statements in a second language (L2) versus the first language (L1) on core knowledge confusion (CKC), superstition, and conspiracy beliefs. Previous research on the Foreign Language Effect (FLE) suggests that using an L2 elicits less intense emotional reactions, promotes rational decision-making, reduces risk aversion, causality bias and superstition alters the perception of dishonesty and crime, and increases tolerance of ambiguity. Our results do not support the expected FLE and found instead an effect of L2 proficiency: Participants with lower proficiency exhibited more CKC, were more superstitious and believed more in conspiracy theories, regardless of whether they were tested in L1 or L2. The study emphasises the importance of considering L2 proficiency when investigating the effect of language on decision-making and judgements: It—or related factors—may influence how material is judged, contributing to the FLE, or even creating an artificial effect.

Highlight

- No evidence supporting the expected Foreign Language Effect (FLE).
- L2 proficiency impacts Core Knowledge Confusion (CKC), superstition, and conspiracy.
- No impact on language used (L1 or L2), L2 proficiency is the key factor.
- Lower L2 proficiency is linked to higher CKC, superstition and conspiracy beliefs.
- L2 proficiency should be considered when studying language effects on judgement.

1. Introduction

1.1. The foreign language effect

Research on the Foreign Language Effect (FLE) comparing bilinguals using their native and foreign language suggests that using a foreign language can influence moral judgement and decision-making (for a review, see: Purpuri, Vasta, Filippi, Wei & Mulatti, 2024). It has been observed that a foreign language elicits less intense emotional reactions than a native language, leading to more rational and utilitarian choices (Corey & Costa, 2015). The use of a foreign language may reduce risk aversion and increase willingness to accept harm for maximising outcomes (Hadjichristidis, Geipel, & Savadori, 2015; Keysar, Hayakawa, & An, 2012; Winkler & Bhatt, 2020; Xing, 2021). Additionally, it has been shown to reduce causality bias (Díaz-Lago & Matute, 2019), suppress superstition (Hadjichristidis, Geipel & Surian, 2019) and alter perceptions of dishonesty and crime severity (Alempaki, Dögan, & Yang, 2020; Woumans, Van der Cruyssen, & Duyck, 2020) with bilinguals perceiving them as less inappropriate and severe in their foreign language. Furthermore, recent evidence suggests that processing information in a foreign language influences the personality trait of tolerance of ambiguity, showing that participants are more tolerant to ambiguity in their foreign language than in their native language (Purpuri, Vasta, Filippi, Wei, & Mulatti, 2023). However, when Vives, Aparici and Costa (2018) investigated the influence of language context on decision-making, across their six studies, it was found that foreign language context did not appear to alter susceptibility to outcome bias or the utilisation of the representativeness heuristic. Additionally, even when emotions were introduced, foreign language context did not significantly influence decision-making tendencies.

Costa, Corey, Hayakawa, Aparici, Vives, & Keysar (2019) investigated the origin of the FLE on moral judgements by assessing whether language context alters the weight given to intentions and outcomes during moral judgement. In his two studies, although the foreign language context reduced the impact of intentions on damage assessment, the overall effect of

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

intention and outcomes on these variables was mainly the same in the foreign and the native language contexts.

1.2. Core knowledge confusion (CKC)

According to the psychological theory of Core Knowledge, infants are born with innate “core knowledge systems” that form the basis for their understanding of the world (Spelke & Kinzler, 2007). These systems encompass four domains: objects, actions, numbers, and space (Spelke & Kinzler, 2007), and provide knowledge beyond the information available in the environment in a universal manner (Hespos & vanMarle, 2012). While people universally expect unsupported objects to fall and understand that hidden objects still exist, confusion or category errors can occur in relation to ontological distinctions, especially concerning mental phenomena, material objects, and living organisms (Lindeman et al., 2011).

The CKC test, developed by Lindeman, assesses these errors, particularly focusing on phenomena associated with paranormal beliefs. In a recent study, Lindeman et al. (2015) identified 14 statements from the test that strongly predicted supernatural beliefs, addressing associations between properties and various ontological categories, typically learned during preschool years (Rakison & Poulin-Dubois, 2001, Wellman & Estes, 1986). These 14 statements from the core knowledge confusions scale ($\alpha = .85$, Lindeman & Aarnio, 2007) concerned the kinds of properties that children learn to associate appropriately with mental, physical, animate, inanimate, living and lifeless phenomena during the preschool years (Rakison & Poulin-Dubois, 2001, Wellman & Estes, 1986). Core ontological confusion emerged as the strongest predictor, supporting Lindeman’s hypothesis that biases in understanding other’s minds contribute to supernatural beliefs. According to Lindeman (Lindeman et al., 2011), confusion and mixing of core ontologies can be explained by dual-process theories (Chaiken & Trope, 1999; Sherman, Gawronski, & Trope, 2014). These suggest that intuitive thinking is automatic and nonverbalizable, relying on associative processes that bypass working memory and conscious attention. In contrast, the analytical system, instead, is specialised for abstract and conceptual problem-solving, involving intentional, verbalizable, and conscious reasoning based on logic and evidence. The two systems usually interact in an integrated manner but can sometimes conflict, allowing for the coexistence of incompatible beliefs. This conflict is often experienced as a struggle between “heart” and “reason,” with the rational system being more verbally justifiable and the intuitive system operating automatically and resisting logical arguments (Epstein et al., 1996). However, Lindeman’s findings do not imply that participants fail to recognise these distinctions when asked directly. Previous research indicates that core knowledge confusions can coexist with scientifically valid knowledge (Subbotsky, 2001; Lindeman & Saher, 2007), and there are two processing modes (Evans, 2008). Lindeman posited that the core knowledge confusions observed in her experiments were more akin to intuitions based on automatic processes, rather than consciously reasoned beliefs.

2. The present study

The main aim of the present study is to investigate the potential L2 effect on CKC. Previous research suggests that using a second language reduces emotional response and promotes rationality (e.g., Costa et al., 2014, Iacozza et al., 2017). Consistently, it has been shown that using a second language suppresses superstition

(Hadjichristidis et al., 2019). Given that CKC correlates with – and might actually cause – superstitious thinking (Lindeman et al., 2011, 2015), it is reasonable to expect an effect of the language used (L1 versus L2) on a test assessing CKC. Specifically, we anticipate that participants responding in L2 will score lower (recognising the real meaning of the items and answering that they are not literally true) on the CKC test than participants responding in L1. We also expect that the language used will not affect participants’ performance on items unrelated to CKC (i.e., genuinely metaphorical items – see materials).

To further explore the rationality aspect, which is indirectly measured through CKC, we included a direct test using three scenarios from Withson and Galinsky (2008), known to be associated with superstition and conspiracy theories. We predicted that participants responding in L1 would exhibit a higher inclination toward belief in superstition and conspiracy theories than those responding in L2 (Corey & Costa, 2015; Hadjichristidis et al., 2019).

Regarding the potential modulatory effect of L2 proficiency, the literature reports mixed results, with some studies finding effects and others not (for a review, see Del Maschio et al., 2022; Pavlenko, 2017; Purpuri, Vasta, Filippi, Wei, Mulatti, 2024). Therefore, we did not expect proficiency in L2 to have any effect on the participants’ responses. However, if we were to make a prediction, participants with higher L2 proficiency responding in L2 would likely perform more similarly to participants responding in L1.

3. Method

3.1. Participants

The sample size was determined using G*power with a between-participants design, $d = .4$ (small/moderate size; c.f. Circi, Gatti, Russo, & Vecchi, 2021), $\alpha = .05$, and power = .8. The analysis indicated a minimum sample size of 156 to detect a between-participant effect of language, with half of the participants (78) being assigned to one condition (L1) and the other half (78) to the second condition (L2) (see pre-registration: <https://osf.io/d4y5p>). Nevertheless, to maximise the study’s statistical power and account for potential drop-outs, we opted to gather data from as many participants as feasible.

A total of 465 participants volunteered for the study. After excluding incomplete responses, data from 304 participants were included in the analysis (195 female, 21 did not provide this information); mean age = 30.2 years (20 participants did not provide this information, age range: 18–70 years). Out of the total participants, 168 were randomly assigned to the Italian questionnaire (L1), and 136 to the questionnaire in English (L2). All participants had Italian as their first language and had studied English as a second language as part of their previous scholastic education. On average, participants in the L2_English condition have had English education since the age of 9.66, with speaking in L2: 9.68 (62 answers not given) – reading in L2: 9.5 (62 answers not given) – writing in L2: 9.79 (62 answers not given).

Participants were asked to self-assess their foreign language proficiency in terms of speaking, reading, writing and understanding levels on a 7-point Likert scale (1 = no competence through 7 = high/native competence). Across the 7 measures, participants in their foreign language condition rated their foreign language skills between 5 and 6 on the Likert scale (English: $M = 5.29$), with speaking at 5.14, reading at 5.72 and writing at 5.03.

Participants were also asked to self-rate their level of English according to the Common European Framework (see Appendix C

Table 1. L2 descriptives table. Only participants in the foreign language condition (L2) are considered

Average age (in years) at which participants started learning L2.			
General	Speaking	Reading	Writing
9.66	9.68	9.5	9.79
Average L2 proficiency			
General	Speaking	Reading	Writing
5.29	5.14	5.72	5.03
Self-rated level of L2 according to the Common European Framework, number of responses (96 missing)			
B1	B2	C1	C2
54	56	75	23

for CEFR descriptors scales of linguistic competence). They could choose between B1 (independent user), B2 (advanced independent user), C1 (proficient user) and C2 (advanced proficient user). Participants who answered the questionnaire in the foreign language condition reported, on average, a level between B2 and C1 (mean_CFR: 2.32 (between b2 and c1; see Table 1). Only 208 participants completed this particular section of the questionnaire. 96 participants did not report this information.

The study (Ref. 2022-028) received ethical approval from the Ethics Committee of the University of Trento and participants provided their consent at the beginning of the online questionnaire.

3.2. Materials

Core Knowledge Confusion. Participants completed an adapted version of the original CKC scale developed by Lindeman et al., 2011 to assess the extent to which university students confuse core ontological attributes related to lifeless material objects, living organisms and mental states. The CKC items that have been used were selected from Lindeman et al. (2015), as they were found to be the most predictive of paranormal beliefs (see Table 2). Filler items, including metaphorical meaning and literal statements, were also included alongside the CKC items (see Appendix A).

Lindeman and colleagues (Lindeman & Aarnio, 2007; Lindeman et al., 2008, 2012, 2015; Svedholm & Lindeman, 2013) grouped the original complete set of statements in the Core Knowledge Confusion Scale into six dimensions, which were labelled as follows: (1) Natural, lifeless entities are living; (2) Force has the attributes of an animate organism; (3) Lifeless entities have the attributes of animate organisms; (4) Inanimate organisms have the attributes of animate organisms; (5) Artificial entities have the attributes of animate organisms; (6) Mental states have the attributes of a physical object.

The results were analysed following Lindeman et al.'s categorization into six dimensions. The six dimensions together with the corresponding items are presented in Table 2 below.

Lindeman et al.'s original materials used in the current study were in English. They were translated into Italian by one of the authors (S.P.). One independent judge controlled the translated version for consistency with the English version. The language versions were also closely matched for word count (English questionnaire = 82 words, Italian questionnaire = 101 words).

Theories of superstition and conspiracy. Participants read scenarios from Studies 3 and 4 of Whitson and Galinsky's studies on

Table 2. List of items sorted by sum variable, following Lindeman et al.'s *Core Knowledge Confusion Scale* (Lindeman et al., 2011)

Dimensions	CKC items
1. Natural, lifeless entities are living	– Rock lives long
2. Force has the attributes of an animate organism	– Force lives in the universe – Force aims to influence
3. Lifeless entities have the attributes of animate organisms	– Planets know things – Earth wants water – The sky hears the thunder
4. Inanimate organisms have the attributes of animate organisms	– Flowers want light – Plants know the seasons
5. Artificial entities have the attributes of animate organisms	– Home knows his residents – Furniture wants a home – House knows its history
6. Mental states have the attributes of a physical object	– Mind touches the other – A plan lives in the nature – Fear poisons man

superstition and conspiracy theories (Whitson, & Galinsky, 2008 – see Appendix A). In Study 3, participants imagined a scenario where they, as successful marketing professionals, forgot their usual pre-meeting ritual and subsequently had their ideas disregarded. They rated the perceived connection between the two events and their level of worry about future performance. In Study 4, participants were presented with two scenarios involving a protagonist experiencing either a bad or a good event, and they rated the extent to which the events were correlated with previously mentioned actions. Ratings were given on a scale from 1 (not at all) to 11 (a great deal).

In the current study, the original English materials were translated into Italian by highly proficient bilingual speakers, the equivalence between the English and the Italian versions was carefully verified through backwards translation and consistency with the English version was ensured by an independent judge.

3.3. Procedure

Participants rated their agreement with the CKC test on a 5-point scale (1 = metaphorical meaning, 5 = literal). They were provided with the following instructions on how to complete the rating of the items: “You are going to read some sentences: some of them have a LITERAL meaning, some of them have a METAPHORICAL meaning. LITERAL sentence example: “Mozart was a composer.” METAPHORICAL sentence example: “A surprising piece of news is a bomb”. Please indicate to what extent the meaning of the following sentences is LITERAL or METAPHORICAL in your opinion.” The presentation order of the 14 CKC test items and the 6 fillers was randomised for each participant. After completing the CKC test, participants were presented with the three superstition and conspiracy scenarios and rated their perceived causal relationship between the described events and the outcome on an 11-point Likert scale (1 = impossible, 11 = certain). Each participant received the questionnaire and the scenarios written entirely in either L1 (Italian) or L2 (English). Following the completion of the questionnaire and the scenarios, participants provided information about their linguistic and cultural background, including self-rating their English proficiency according to the Common European Framework of Reference for Languages (CEFR) descriptor scales (see Appendix C).

4. Results

Data were analysed in two ways. In the *first pre-registered analysis* we looked at the effect of language on CKC using data from all the 304 participants who completed the questionnaire. In the *second explorative analysis*, we considered only data from the 208 participants who rated their proficiency in L2 according to CEFR descriptors. We created a two-level variable by grouping participants according to their Proficiency in L2: participants reporting either B1 or B2 were classified as Low Proficient (110 participants) while participants reporting either C1 or C2 were classified as high-proficient (98 participants).

First analysis. The 14 CKC items were entered as dependent variables in a multivariate analysis of variance (MANOVA), with Language of the Questionnaire (Italian versus English) as the between-participant factor. As expected, this revealed a main effect of Language of the Questionnaire, $F(1, 302) = 6.792$, $p < .001$. To understand this result within a general framework, we chose to average the 14 CKC items into a single variable and analyse them by means of an analysis of variance (ANOVA) with Language of the Questionnaire (Italian versus English) as the between-participants factor. In contrast to our predictions, this indicated a significant effect of Language of the Questionnaire, $F(1, 302) = 7.0$, $p = .009$, $\eta_p^2 = .023$, with participants who completed the questionnaire in English (mean = 2.56, $sd = 0.603$) scoring higher (i.e., assigning a greater literal meaning to the items) than participants who completed the questionnaire in Italian (mean = 2.37, $sd = 0.629$). Surprisingly, this outcome suggests that the use of L2 might lead to increased CKC scores rather than a decrease as initially hypothesised.

Mean scores for superstition and conspiracy scenarios were analysed by means of an ANOVA with the Language of the Questionnaire as a between-participant factor. The Language of the Questionnaire exerted no significant effects, $F < 1$.

Second analysis. To analyse CKC scores an ANOVA was performed with Language of the Questionnaire (Italian versus English) and Proficiency in L2 (High Proficient versus Low Proficient) as between-participants factors, and Item Type (CKC items versus Metaphorical fillers) as within-participant factor. The ANOVA showed: a main effect of Language of the Questionnaire, $F(1, 204) = 14.2$, $p < .001$, $\eta_p^2 = .065$, thus confirming the previous analysis; a main effect of Proficiency in L2, $F(1, 204) = 12.9$, $p < .001$, $\eta_p^2 = .060$, with low-proficient participants scoring higher than high-proficient participants; and a main effect of Item Type, $F(1, 204) = 151.2$, $p < .001$, $\eta_p^2 = .426$, with scores to CKC items being higher than scores to Metaphorical fillers. In addition, the ANOVA showed a significant interaction both between Language of the Questionnaire and Proficiency in L2, $F(1, 204) = 11.1$, $p = .001$, $\eta_p^2 = .052$, and between Language of the Questionnaire and Item Type, $F(1, 204) = 6.5$, $p = .011$, $\eta_p^2 = .031$. Interestingly, the ANOVA also showed a significant three-way interaction, $F(1, 204) = 10.3$, $p = .002$, $\eta_p^2 = .048$. This three-way interaction was explored with interactive contrasts, which showed that such interaction was driven by the fact that Language of the Questionnaire significantly interacted with Proficiency in L2 for Metaphorical fillers (Figure 1, right panel), $F(1, 204) = 17.9$, $p < .001$, $\eta_p^2 = .081$, whereas the two variables did not significantly interact for CKC items (Figure 1, left panel), $F(1, 204) = 1.5$, $p = .40$. In other words, proficiency in L2 affected responding to Metaphorical fillers (i.e., unambiguously metaphorical items) written in English (L2), with low-proficient participants giving higher scores (i.e., less accurate judgements) than high-proficient participants, but not to Metaphorical fillers written in Italian (L1), to which both low and high-proficient participants gave low scores. In contrast, being low or high proficient in L2 affected responding to CKC items both when the items were presented in English (L2) and when they were presented in Italian

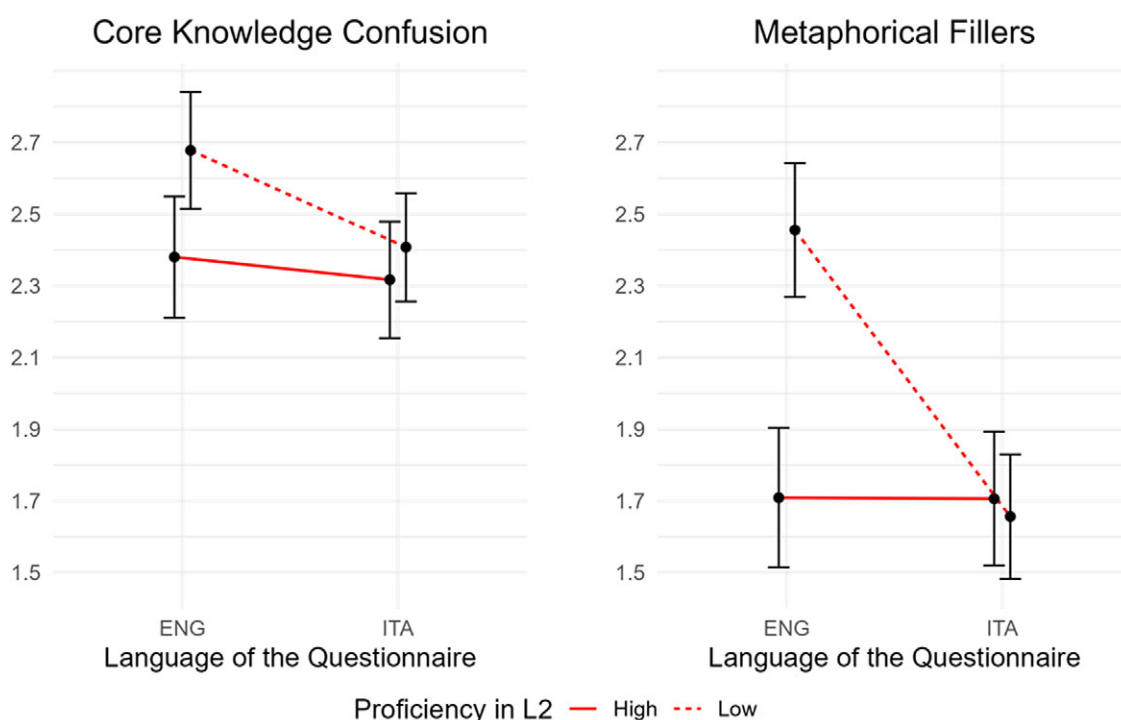


Figure 1. Mean CKC (left) and metaphorical filler (right) scores were sorted by language of the questionnaire and proficiency in L2. Vertical bars represent 95% confidence intervals.

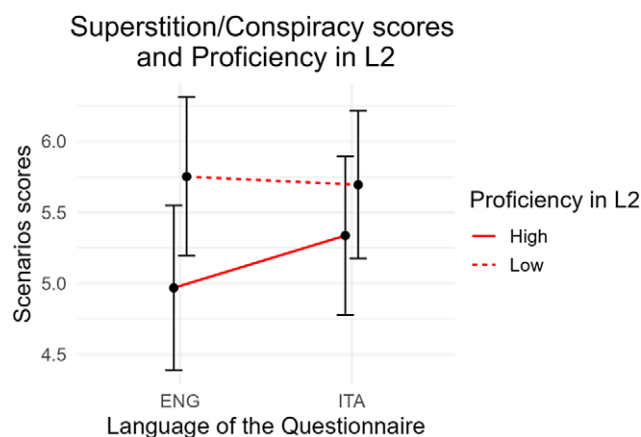


Figure 2. Aggregated scores of all scenarios, sorted by language of the questionnaire and proficiency in L2. Vertical bars represent 95% confidence intervals.

(L1): in both cases, low-proficient participants gave lower scores than high proficient participants.

Mean scores for superstition and conspiracy scenarios were analysed by means of an ANOVA with the Language of the Questionnaire and Proficiency in L2 as between-participants factors. Whereas the effect of Proficiency in L2 proved significant, $F(1, 204) = 4.1$, $p = .044$, $\eta_p^2 = .020$, neither the main effect of Language of the Questionnaire nor the interaction was significant, $F_s < 1$. As Figure 2 shows, that proficiency in L2 affects participants' performance regardless of whether they are responding to scenarios presented in L1 or L2, with High Proficiency participants giving less superstitious/conspiracy judgments (i.e., lower scores) than Low Proficiency participants overall.

5. Discussion

The present study investigated whether bilingual individuals exhibit increased rationality and are less prone to believe in superstitious beliefs when using their second language (L2). If this hypothesis holds true, we would anticipate observing a heightened level of rationality in participants' responses to the CKC test and the scenarios related to superstition and conspiracy when administered in their L2.

We assessed 304 bilingual speakers through an online questionnaire in either English or Italian, measuring CKC as well as superstition and conspiracy theories. Participants judged the metaphorical or literal nature of CKC items and the connection between superstition and conspiracy scenarios and the given outcomes.

The analyses of CKC scores revealed a language effect, with participants who completed the questionnaire in English exhibiting higher scores compared to those completing it in their first language. This result contradicted our predictions and previous research, suggesting that the use of L2 increased CKC scores overall. Proficiency in L2 and item type also showed main effects. The item-type effect was an expected one: truly metaphorical filler items were judged as more metaphorical (i.e., less literal) than CKC items, thus replicating Lindemann's findings. Notably, a significant three-way interaction indicated that L2 proficiency affected responses to CKC items regardless of language, while its impact on metaphorical fillers was evident only in the L2 condition. Contrary to expectations, language did not significantly affect scores on superstition

and conspiracy scenarios. However, proficiency in L2 did influence the scores, with low-proficient participants scoring higher than high-proficient participants, irrespective of the questionnaire language. Participants with higher English proficiency demonstrated less superstition and belief in conspiracy theories.

This suggests that individuals with lower L2 proficiency encounter challenges in understanding and responding to CKC items, irrespective of the language in which they are presented, and the impact of L2 proficiency on CKC judgements seems to transcend the specific language used. However, when it comes to metaphorical filler items that are clearly metaphorical, the influence of L2 proficiency was evident only in the L2 condition. In the L1 condition, both low and high-proficient participants tended to judge the metaphorical filler items accurately, whereas in the L2 condition, low-proficient participants were less accurate than high-proficient participants. Therefore, the impact of L2 proficiency on the ability to judge metaphorical items seems to transcend the specific language used, but only when the metaphorical meaning is not straightforward.

Contrary to our expectations, the questionnaire results related to superstition and conspiracy beliefs did not show significant effects of language, indicating that language did not affect participants' scores in these scenarios. Therefore, this study did not replicate previous findings on this issue (Hadjichristidis et al., 2019; Pavlenko, 2017; see also Del Maschio et al., 2022). However, an effect of proficiency in L2 on the scenarios' scores was found, regardless of the language used in the questionnaire: Interestingly, this effect did not interact with the Language of the Questionnaire itself. Participants who were more proficient in English, regardless of whether they responded in their native or second language, showed lower levels of superstitious beliefs and were less prone to believing in conspiracy theories. The impact of L2 proficiency on beliefs has yielded mixed findings in the literature, with some studies reporting significant effects of L2 proficiency on beliefs (Costa et al., 2014; Hadjichristidis et al., 2019, Study 1), while others have not (Hadjichristidis et al., 2019, Study 2; Geipel et al., 2015b, Study 3). It is important to note that in these studies, the effect of proficiency in L2 was assessed only for participants responding in L2 and not for participants responding in L1. In our study, L2 proficiency affected both responses in L2 and in L1.

To understand the significance of these preliminary findings, it is crucial to interpret and consider their implications. If we assume that the FLE reflects heightened rationality when responding in L2 compared to L1, we would expect that participants in this study score lower in both the CKC scale and the scenarios when responding in L2 compared to L1. However, our results show the opposite effect: higher scores on the CKC scale (i.e., an opposite FLE) and no discernible effects on the scenarios. Although further studies are needed to confirm this result, these findings clearly cast doubts on the FLE phenomenon and its conventional interpretation.

Nevertheless, it is important not to dismiss the perspective of the FLE phenomenon entirely. Instead, we should explore the implications of the observed effect of L2 proficiency. First, it can be argued that as participants become more proficient in L2, their responses in L2 should become more similar to those of participants responding in L1. Our findings support this argument. Second, if when responding in a more proficient language people exhibit less rationality, then highly L2-proficient participants should show lower levels of rationality (i.e., higher scores on both the CKC scale and the scenarios) compared to low-proficient participants when

responding in L2. However, both the CKC and the scenario results contradict this expectation. Low-proficient participants attributed higher “literal meaning” scores than high-proficient participants, especially when responding in L2. In summary, predictions based on the notion of the FLE as a variation along the dimension of rationality conflict with our findings.

We propose an alternative explanation. Let us examine the results pertaining to the straightforward metaphorical filler items (Figure 1, right panel). Participants who responded in Italian accurately classified these items as genuinely metaphorical, with most responses falling within the range of 1.5 to 1.9 on a scale of 5. Highly L2 proficient participants also showed similar results with the majority of responses falling in the 1.5 to 1.9 range. In other words, both Italian participants responding in Italian and highly proficient Italian participants responding in English correctly identified clearly metaphorical items. However, low-proficient Italian participants responding in English were unable to classify the metaphorical filler items correctly, as their responses mostly fell within the 2.3–2.7 range. We suggest that this particular group of participants lacks the ability to accurately process English sentences, which leads them to treat even unambiguous sentences as ambiguous. Less proficient participants may be less confident while processing English materials, and this uncertainty influences their judgments.

Let us now focus on the CKC scale (Figure 1, left panel) and analyse the data based on our proposed perspective. The CKC items (e.g., “Planets know things”) are more ambiguous in terms of their metaphorical meaning compared to the metaphorical filler items (e.g., “Howling wind is a flute”), resulting in higher scores that indicate a more literal interpretation (see Lindemann, 2011). Similar to the metaphorical fillers, it is plausible that low-proficient participants responding in English exhibited less confidence in their sentence interpretations compared to highly proficient participants. This increased uncertainty among low-proficient participants is reflected in the scores they assign, which tend to gravitate towards the middle of the 5-point scale. This interpretation can also be extended to the results obtained from the scenarios related to superstition and conspiracy theories. However, even when responding in Italian, low-proficient participants attributed higher scores to CKC items and superstition and conspiracy scenarios compared to highly proficient participants. In other words, L2 proficiency impacts not only performance in L2 but also performance in L1.

Therefore, the main findings of this study show for the first time the impact of L2 proficiency on the ability to judge ambiguous sentences or scenarios. These findings cannot be explained by the standard interpretation of the FLE, as this interpretation fails to account for the effect of L2 proficiency observed when participants judged material both written in English (L2; as explained above, this account would predict less rational judgement for high than low-proficient participants) and written in Italian (L1). There is no reason to expect that highly L2-proficient participants would exhibit either greater or lesser emotional distance from Italian stimuli (Purpuri et al., 2024).

Is there some kind of factor associated with proficiency in L2 that hinders rational judgement of ambiguous sentences or event-consequence scenarios even when presented in L1? This study shows that this seems to be the case. We propose that this factor is not proficiency itself, but rather some unknown variable that correlates with proficiency (see, for example, de Bruin, Treccani, & Della Sala, 2015; Treccani & Mulatti, 2015). Although we are unsure about the nature of this variable (or set of variables), we

can exclude some candidates, such as years of instruction, age, and gender, as these variables were balanced between the highly proficient and low-proficient groups in our study ($F_s < 1$). Instead, we suggest some possible (linguistic and extralinguistic) factors, such as language comprehension skills, general knowledge, educational background, and cultural exposure (see Purpuri et al., 2024), individual interests or opportunities, fluid intelligence (gF) and/or domain-general executive functioning (EF).

These last two factors are of particular interest and are fairly likely candidates. Fluid intelligence and EF are critical cognitive abilities that support complex reasoning and decision-making processes, and their variability among individuals may significantly impact both L2 proficiency and performance on a variety of tasks in L1, as well as in L2. Specifically, lower levels of gF and EF could lead to both reduced L2 proficiency and difficulties with tasks requiring the discrimination of nuanced language, such as distinguishing between literal and metaphorical statements, regardless of the language of these statements, particularly when they are ambiguous. In the case of unambiguous statements (i.e., the filler CKC items in the present study), even relatively low levels of gF and EF might allow for correct judgments of the literal versus metaphorical meaning of the statements, at least when they are presented in one’s native language. Moreover, these cognitive factors could also contribute to a higher tendency among individuals with lower L2 proficiency to endorse superstitions and conspiracies. Accordingly, the observed relationship between L2 proficiency and the endorsement of such beliefs in both L1 and L2 (see Figure 2) might be mediated by gF and EF rather than being a direct effect of L2 proficiency alone. Given the potential for wide individual differences in gF and EF, even in a sample where the age of acquisition (AoA) for L2 is relatively consistent, it is plausible that these cognitive abilities could drive the observed differences in both L2 proficiency and performance on the CKC and Superstition/Conspiracies tasks. To address this potential confound in future research, it would be valuable to include assessments of gF and EF, (e.g., through abstract reasoning tests, such as Raven’s (2003) Progressive Matrices, and response conflict tests, such as the Stroop task; Burgoyne et al., 2023). This assessment could help clarify the extent to which these cognitive factors contribute to the patterns observed in the current study.¹

It is worth reiterating that the L2 proficiency effect we observed in our study when the material was in L1 (thus, not attributable to differences in confidence when interpreting sentences written in a foreign language) is specific to the ambiguity of the material. This effect is evident with the CKC items (which are ambiguous along the metaphorical-literal continuum) and the scenarios (which lack a straightforward response and are inherently ambiguous), but proficiency in L2 does not influence the response to clearly metaphorical filler items in Italian. This could shed new light on the results of previous studies that found an impact of the ambiguity of the material used on the FLE.

Costa et al. (2014) previously argued that a certain level of ambiguity is necessary for the FLE to manifest. Building on the results of their first experiment, which were critically tied to the ambiguity or uncertainty of the material used, they conducted a second experiment that involved both ambiguous and

¹We thank Kenneth Paap for raising this important point in his review of the paper.

unambiguous material. In this second experiment, they found the FLE only with the ambiguous material. This observation, along with our results, leads us to consider the intriguing possibility that the L2 proficiency of the participants in this study—or rather, the linguistic or extralinguistic factors associated with it—may have impacted how they interpreted L1 and L2 items and contributed to the FLE observed by these authors.

In conclusion, our results challenge the conventional interpretation of the FLE and suggest that the observed impact of language on emotional reactions, decision-making, rationality, moral judgement, and cognitive biases is intricately linked to participants' L2 proficiency. This raises concerns regarding previous studies that have exclusively focused on the FLE, as they may have overlooked the crucial role of participants' L2 proficiency. It is possible that variations in the FLE in CKC could be attributed to participants not fully understanding the items. Therefore, the existing literature should probably be reevaluated considering the possibility that it is L2 proficiency that affects the results, rather than the language of presentation. For instance, if participants who judged the material in L2 had different (higher) L2 proficiency compared to those who judged it in L1, this could lead to an artificial FLE.

Our findings underscore the significance of considering participants' L2 proficiency. By recognizing it, future research can provide a more nuanced understanding of its role. Additionally, these results emphasise the importance of properly assessing and accounting for participants' language skills in experimental designs involving a foreign language context.

References

- Alempaki, D., Doğan, G., & Yang, Y. (2020). Lying in a foreign language? Retrieved from SSRN: DOI:10.2139/ssrn.2995506
- Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General*, 152(8), 2369.
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.
- 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. DOI:10.3758/s13423-020-01871-z
- Corey, J. D., & Costa, A. (2015). El efecto de usar una lengua extranjera sobre las decisiones morales. *Ciencia Cognitiva*, 9(3), 53–56.
- Costa, A., Corey, J. D., Hayakawa, S., Aparici, M., Vives, M.-L., & Keysar, B. (2019). The role of intentions and outcomes in the foreign language effect on moral judgements. *Quarterly Journal of Experimental Psychology*, 72(1), 8–17. DOI:10.1177/1747021817738409
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguía, J., Heafner, J., et al. (2014). Your morals depend on language. *PLoS ONE*, 9(4), e94842. DOI: 10.1371/journal.pone.0094842
- De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias?. *Psychological Science*, 26(1), 99–107. DOI:10.1177/0956797614557866
- 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. DOI: 10.1017/S1366728921001012
- Díaz-Lago, M., & Matute, H. (2019). Thinking in a foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, 72(1), 41–51.
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, 71(2), 390.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, 59, 8–17. DOI:10.1016/j.jesp.2015.02.001.
- Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied*, 21(2), 117–129. DOI:10.1037/xap0000044
- Hadjichristidis, C., Geipel, J., & Surian, L. (2019). Breaking magic: Foreign language suppresses superstition. *Quarterly Journal of Experimental Psychology*, 72(1), 18–28. DOI:10.1080/17470218.2017.1371780
- Hespos, S. J., & vanMarle, K. (2012). Physics for infants: Characterizing the origins of knowledge about objects, substances, and number. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(1), 19–27.
- Iacozza S., Costa A., Duñabeitia JA (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PLoS ONE*12(10), e0186027. 10.1371/journal.pone.0186027
- 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. DOI:10.1177/0956797611432178
- Lindeman, M., & Aarnio, K. (2007). Superstitious, magical, and paranormal beliefs: An integrative model. *Journal of Research in Personality*, 41(4), 731–744. <https://doi.org/10.1016/j.jrp.2006.06.009>
- Lindeman, M., Blomqvist, S., & Takada, M. (2012). Distinguishing spirituality from other constructs: Not a matter of well-being but of belief in supernatural spirits. *Journal of Nervous and Mental Disease*, 200(2), 167–173. <https://doi.org/10.1097/NMD.0b013e318243953e>
- Lindeman, M., Cederström, S., Simola, P., Simula, A., Ollikainen, S., & Riekk, T. (2008). Sentences with core knowledge violations increase the size of N400 among paranormal believers. *Cortex*, 44(10), 1307–1315. <https://doi.org/10.1016/j.cortex.2007.07.010>
- Lindeman, M., & Saher, M. (2007). Vitalism, purpose and superstition. *British Journal of Psychology*, 98(1), 33–44.
- Lindeman, M., Svedholm-Häkkinen, A., & Lipsanen, J. (2015). Ontological confusions but not mentalizing abilities predict religious belief, paranormal belief, and belief in supernatural purpose. *Cognition*, 134, 63–76.
- Lindeman, M., Svedholm-Häkkinen, A., Takada, M., Lönnqvist, J.-E., & Verkasalo, M. (2011). Core knowledge confusions among university students. *Science & Education*, 20, 439–451. 10.1007/s11191-009-9210-x.
- Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current Opinion in Psychology*, 17, 74–78. DOI: 10.1016/j.copsyc.2017.06.005
- Purpuri, S., Mulatti, C., Filippi, R., & Treccani, B. (2024). Inside the kaleidoscope: Unravelling the “feeling different” experience of bicultural bilinguals. *Frontiers in Psychology*, 15, 1376076.
- Purpuri, S., Vasta, N., Filippi, R., Wei, L., & Mulatti, C. (2023). The foreign language effect on tolerance of ambiguity. *Bilingualism: Language and Cognition*, 1–9.
- Purpuri, S., Vasta, N., Filippi, R., Wei, L., & Mulatti, C. (2024). Does language shape the way we think? A review of the foreign language effect across domains. *International Journal of Bilingualism*, 2024. DOI: 10.1177/13670069231225374
- Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, 127(2), 209–228. DOI: 10.1037/0033-2909.127.2.209
- Raven, J. (2003). Raven progressive matrices. In: *Handbook of nonverbal assessment* (pp. 223–237). Boston, MA: Springer US.
- Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. Guilford Publications.
- Svedholm, A., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology*, 104(3), 303–319. <https://doi.org/10.1111/bjop.12031>
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89–96. DOI: 10.1111/j.1467-7687.2007.00569.x
- Subbotsky, E. (2001). Causal explanations of events by children and adults: Can alternative causal modes coexist in one mind?. *British Journal of Developmental Psychology*, 19(1), 23–45.
- Treccani, B., & Mulatti, C. (2015). No matter who, no matter how... and no matter whether the white matter matters. Why theories of bilingual

- advantage in executive functioning are so difficult to falsify. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, **73**, 349–351. DOI: [10.1016/j.cortex.2015.07.015](https://doi.org/10.1016/j.cortex.2015.07.015)
- 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.
- Wellman, H. M., & Estes, D.** (1986). Early understanding of mental entities: A reexamination of childhood realism. *Child Development*, **57**(4), 910–923. DOI: [10.2307/1130367](https://doi.org/10.2307/1130367)
- Whitson, J. A., & Galinsky, A. D.** (2008). Lacking control increases illusory pattern perception. *Science*, **322**(5898), 115–117. DOI: [10.1126/science.1159845](https://doi.org/10.1126/science.1159845)
- Woumans, E., Van der Cruyssen, I., & Duyck, W.** (2020). Crime and punishment: Morality judgment in a foreign language. *Journal of Experimental Psychology: General*, **149**(8), 1597–1602. DOI: [10.1037/xge0000736](https://doi.org/10.1037/xge0000736)
- Xing, A.** (2021). Foreign language effect on risk preference: The framing effect, loss aversion, and risk aversion. *Journal of Behavioral and Experimental Economics*, **93**, 101723. DOI: [10.1016/j.socec.2021.101723](https://doi.org/10.1016/j.socec.2021.101723).