Bilingualism: Language and Cognition

cambridge.org/bil

Boundaries of the foreign language effect under conditions of uncertainty

Nicola Del Maschio[1], Camilla Bellini[1], Jubin Abutalebi[1,2] and Simone Sulpizio[3,4]

[1]Centre for Neurolinguistics and Psycholinguistics, Faculty of Psychology, Vita-Salute San Raffaele University, Milan, Italy; [2]Psycholinguistics of Language Representation (PoLaR) Lab, Center for Language, Brain and Learning (C-LaBL), UiT-The Arctic University of Norway, Tromsø, Norway; [3]Department of Psychology, University of Milano-Bicocca, Milan, Italy and [4]Milan Center for Neuroscience (NeuroMI), University of Milano-Bicocca, Milan, Italy

Research Article

Cite this article: Del Maschio, N., Bellini, C., Abutalebi, J. and Sulpizio, S. (2026). Boundaries of the foreign language effect under conditions of uncertainty. *Bilingualism: Language and Cognition* **29**, 784–794. https://doi.org/10.1017/S1366728925100400

Received: 2 May 2024
Revised: 23 June 2025
Accepted: 10 July 2025
First published online: 19 August 2025

Keywords:
foreign language effect; bilingualism; decision-making; uncertainty; individual differences

Corresponding author:
Nicola Del Maschio;
Email: ndelmaschio@gmail.com

This research article was awarded Open Data badges for transparent practices. See the Data Availability Statement for details.

Abstract

People's decisions may change when made in a foreign language (FL). Research testing this foreign language effect (FLE) has mostly used scenarios where uncertainty is expunged or reduced to a form of risk, whereas real-life decisions are usually characterized by uncertainty around outcome likelihood. In the current work, we aimed to investigate whether the FLE on decision-making extends to uncertain scenarios. Moreover, as it is still unclear what linguistic and psychological factors contribute to the FLE, we tested the effects of participants' FL background, cognitive style and risk-taking attitude on decision processes under certain and uncertain conditions. Overall, we report null effects of language context (native versus foreign language) and problem condition (certain versus uncertain prospects) on participants' choices. In addition, we found that both FL background and decision makers' traits modulated participants' choices in a FL, without emerging into the 'classic' FLE on decision-making. However, the direction of such effects was complex, and not always compatible with previous FLE theories. In light of these results, our study highlights the need to reconceptualize the FLE and its implications on decision-making.

Highlights

- We investigated whether the FLE on decision-making extends to uncertain scenarios.
- We explored the impact of linguistic and psychological factors on bilinguals' choices.
- We report null effects of language context and problem condition on decision-making.
- Both FL background and decision makers' traits modulated participants' choices in a FL.
- The direction of such effects was complex and often incompatible with previous theories.
- Our results call for a general rethinking of the FLE and its underlying mechanisms.

1. Introduction

Our ability to make effective decisions is affected by the quality of information we receive, and such information is often communicated linguistically (Costa et al., 2017; Keller & Staelin, 1987). In recent years, a growing body of literature has shown that language and decision-making interact in guiding our choices, and, perhaps more strikingly, that decision outcomes may change when information is presented in either our native (NL) or a foreign language (FL) (see, for meta-analyses, Del Maschio, Crespi, et al., 2022a; Purpuri et al., 2024a). This language-of-presentation effect – known as the so-called 'Foreign Language Effect (FLE)' – has been found in multiple domains of decision-making, from people's treatment of risk, losses and gains (e.g., Costa, Foucart, Arnon, et al., 2014a) to superstitious thinking (Hadjichristidis et al., 2019) and moral judgment (e.g., Costa, Foucart, Hayakawa, et al., 2014b). One key result of this stream of research is that processing information in a FL would reduce susceptibility to cognitive biases, possibly by affecting the contributions of intuition and deliberative reasoning in our decision-making processes. It has been shown, for instance, that individuals operating in a FL are more likely to display reduced risk-aversion in monetary gambles with positive expected value (e.g., Costa, Foucart, Arnon, et al., 2014a) and increased utilitarian behavior in moral dilemmas tailored to the deontological/utilitarian dichotomy (e.g., Cipolletti et al., 2016). Moreover, evidence has been provided that using a FL may diminish the tendency to perceive causal relationships between unrelated events (Díaz-Lago & Matute, 2019), reduce common superstitious beliefs (Hadjichristidis et al., 2019) and increase tolerance for ambiguity (Purpuri et al., 2024b), dishonesty (Yang et al., 2021) and crime (Woumans et al., 2020). As of yet, however, research on the FLE has mostly focused on two particular classes of decision-making contexts, i.e., those characterized by certainty and known risk. Under conditions of certainty, all relevant alternatives are known, and each action leads to a certain outcome. Certainty is reflected in the typical

© The Author(s), 2025. 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.

structure of moral dilemmas, where the outcomes of actions are presented as completely deterministic (e.g., 'If you push the man on the footbridge, five people will be saved; if you don't, five people will be killed'). Under known risk, all relevant alternatives are known, and the probabilities of all outcomes for each alternative can be computed. The architecture of known risk is reflected in certain types of gambling, such as dice games or lotteries. For example, in a typical paradigm involving risky financial decisions, participants must choose between the sure option of receiving €1 or gamble for a 50% chance to get nothing or €2.50.

Notably, in both moral and risky decision-making contexts, experimental scenarios depict 'small worlds' of perfect information where uncertainty is expunged or reduced to a form of risk, and optimal strategies for utility maximization can be implemented. However, this is rarely the case in everyday experience. In fact, the majority of real-life scenarios we encounter are situations in which the probability of future events and their possible outcomes cannot be expressed with mathematical precision (Mousavi & Gigerenzer, 2014). Put differently, we live in 'large worlds' where some of the alternatives and outcomes, in addition to their probabilities, are not known for certain.

In the current work, we aimed to investigate whether the FLE on decision-making extends to uncertain scenarios, which characterize most of the decision-making contexts we encounter. In addition, as it is still unclear what linguistic and psychological factors contribute to the FLE, a second aim of the current study was to investigate the effects of participants' FL background, cognitive style and risk-taking attitude on decision-making processes under certain and uncertain conditions. By doing this, we expected to provide a more comprehensive and ecological picture of the FLE on choice behavior. Moreover, we expected to gain deeper insights into the psychological underpinnings of this phenomenon, the etiology and characteristics of which remain poorly understood.

1.1. The present study

As highlighted above, research on the FLE has been largely biased toward informational conditions where all risks are known and optimization is possible. The primary aim of the current study was to investigate the FLE in decision-making contexts where there is uncertainty related to potential outcomes. To achieve this goal, we presented to our participants, in either a NL or a FL context, two kinds of problems characterized by uncertainty around outcome likelihood: (1) Moral dilemmas with uncertain prospects, (2) Exploration–exploitation problems. Both kinds of problems entail a trade-off between conflicting objectives. In moral dilemmas, the decision-maker must decide between committing a moral violation in order to maximize overall outcomes (utilitarian option) or rejecting the moral violation based on consistency with moral rules (deontological option). In exploration–exploitation problems, the decision-maker must decide between exploiting a known source of reward (i.e., an option with a known beneficial outcome, but which may devalue over time) or exploring the environment to find alternative sources of reward (i.e., options of which little is known about, but which might turn out to be of superior value). In a between-group analysis, we investigated the effects of language context and problem condition on participants' choices. Across all problems, we also evaluated the perceived emotional distress associated with processing each scenario, a choice motivated by explanatory hypotheses relating the FLE to reduced emotionality in FLs (e.g., Caldwell-Harris, 2015; Costa, Foucart, Hayakawa, et al., 2014b).

Especially, early studies in the FLE literature have suggested that the systematic deviations from norm or rationality that allegedly characterize human reasoning (see Gilovich et al., 2002) are reduced when operating in a FL, possibly reflecting a reduced impact of intuition and/or an increased impact of cognitive control on decision processes (Costa, Foucart, Arnon, et al., 2014a; Keysar et al., 2012). More recently, however, null findings from various studies have cast doubt on the generalizability of the FLE (e.g., Mækelæ & Pfuhl, 2019; Muda, Walker, et al., 2020; Vives et al., 2018), inspiring increasing interest in the mechanisms and contexts modulating its occurrence.

A second aim of the current study was to investigate the effects of participants' FL background, cognitive style and risk-taking attitude on bilinguals' decision-making. To achieve this goal, in a within-group analysis, we focused on participants assigned to the FL context and examined whether their choices and perceived emotional distress were modulated by individual differences in FL background. We extended to the FLE domain a conceptualization of bilingualism that frames the phenomenon as a construct comprising several interrelated dimensions and explored the extent to which bilinguals varied along FL proficiency, FL exposure and FL immersion – all measured as continuous variables (see Del Maschio, Del Mauro, et al., 2022b; Sulpizio et al., 2020). In addition, we tested for potential interactions between significant effects of FL background and decision-makers' traits on choice behavior and distress ratings. The inclusion of these measures allowed us to look into whether the effects of variables related to bilinguals' language experience were influenced by individual dispositions toward risk and reflective thinking.

Based on previous evidence on monolingual (e.g., Kortenkamp & Moore, 2014) and bilingual individuals (e.g., Hadjichristidis et al., 2015), we hypothesized that the FLE would extend to uncertain scenarios. Consistent with Del Maschio, Del Mauro, et al. (2022b), we also assumed that FL proficiency, FL exposure and FL immersion would interact in modulating both decision outcomes and distress ratings in FL contexts. In particular, higher levels of proficiency and exposure to the FL, as well as continuous FL usage in immersive contexts, were expected to reduce utilitarian inclinations in moral dilemmas and exploratory behavior in exploration–exploitation problems. In line with the hypothesis that high levels of proficiency and exposure to the FL promote emotional grounding (see, Hayakawa et al., 2016), increased proficiency and exposure were also expected to generate higher emotional distress associated with FL processing. Earlier work indicated that propensity toward reflective (versus intuitive) thinking can predict utilitarian (versus deontological) moral judgments (e.g., Paxton et al., 2012), and that risk and novelty seeking can bias decisions toward exploration (e.g., Wittmann et al., 2008). Based on these findings, we predicted that significant effects of FL background would be modulated by measures of cognitive reflection and risk-taking attitude.

2. Materials and methods

2.1. Participants

The sample size was established on the basis of previous FLE studies that operationalized bilingualism as a continuous variable (e.g., Del Maschio, Del Mauro, et al., 2022b; Muda, Walker, et al., 2020; Privitera et al., 2023). Three hundred thirty-nine young adults (234 F, M_{Age} = 25, ± SD_{Age} = 4.27 years) volunteered to participate in the study. Participants were recruited via advertisements on university bulletin boards and social media. All participants were native Italian speakers (NL) who spoke English as a foreign language (FL). Participants were randomly assigned to a NL context

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

(Italian; N = 168) and a FL context (English; N = 171). Across language contexts, participants were matched on age (NL: M = 24.82, SD = 4.11; FL: M = 25.19, SD = 4.44; W = 15025, p = .462), gender (NL: 123 F; FL: 111 F, χ^2 = 2.357, p = .125) and education (NL: M = 16.15, SD = 2.17; FL: M = 16.58, SD = 1.97; W = 16004, p = .063). All participants declared no history of neurological or psychiatric disease, nor current treatment with psychiatric medications. All participants gave their informed consent prior to being enrolled. The study was approved by the Human Research Ethics Committee of the San Raffaele Hospital (Milan, Italy).

2.1.1. Foreign language background

Participants' FL background was assessed by means of both subjective and objective measures. The Language History Questionnaire (version 3) (LHQ3; Li et al., 2020) was used to assess participants' FL Age of Acquisition (AoA), Proficiency, Exposure and Immersion. AoA was operationalized as the lowest age at which participants began to listen to, or learn to speak or write, in the FL. Proficiency was calculated as the weighted sum of participants' self-ratings on their level of speaking, listening, reading and writing in the FL. Participants rated their proficiency level on a 7-point scale (1 = 'Very poor'; 7 = 'Excellent'). The amount of Exposure to the FL was computed by asking participants how many hours per day they spent speaking, listening, reading and writing in their languages at the time of testing. A normalized score for FL daily Exposure was then computed (0 = No exposure to the FL; 1 = Exposure to the FL only). FL Immersion was computed as an aggregate score based on participants' current age, FL AoA and years spent using the FL. Moreover, an objective measure of FL Proficiency was collected by means of the online English Language Assessment (ELA) developed by Cambridge (https://www.cambridgeenglish.org) (see Sulpizio et al., 2019). The ELA includes 25 multiple-choice items to assess general English knowledge, requiring participants to choose the correct grammar option to fill in the blanks within English sentences. The proficiency score was calculated upon the sum of correct responses (score range: 0–25 points). The score provides an estimation of English proficiency in terms of the reference levels defined by the Common European Framework of Reference for Languages (i.e., A1, A2, B1, B2, C1, C2). The objective proficiency levels of participants assigned to the FL context were distributed as follows: A2 = 19% (N = 33), B1 = 21% (N = 35), B2 = 26% (N = 45), C1 = 13% (N = 23), C2 = 12% (N = 20). Fifteen participants (9%) did not complete the ELA.

Descriptive statistics of participants' FL background are reported in Table 1.

2.1.2. Cognitive style and risk-taking attitude

Participants' cognitive style was evaluated through the Cognitive Reflection Test (CRT; Frederick, 2005). The CRT measures the propensity to engage in reflective thinking in lieu of prepotent responding. In particular, the CRT presents three mathematical problems that trigger an intuitive but incorrect response, which requires reflective thinking to be ultimately rejected (see Supplementary Material for the full illustration of problems). As such, the CRT is believed to provide an indication of people's ability to 'override' an incorrect impulsive response to counterintuitive problems. The CRT can be scored using different procedures (Erceg & Bubić, 2017). We opted for the most commonly used procedure, that is, the sum of participants' correct responses on the three problems, ranging from 0 to 3. This score is supposed to reflect individual differences in reflective thinking (i.e., the ability to reflect upon, and ultimately override, intuitive responses). The higher the score, the more reflective thinking is employed to solve the problems.

Participants' risk-taking attitude was assessed by means of the Italian adaptation (Fossati & Somma, 2021) of the Personality Inventory for DSM-5 (PID-5) – Adult (Krueger et al., 2012). Only the 14 items pertaining to the personality trait facet of risk-taking were used (i.e., items 3, 7, 35, 39, 48, 67, 69, 87, 98, 112, 159, 164, 195, 215). Each item refers to a statement regarding how people may describe themselves in risky situations. Participants were asked to carefully read and rate how well each statement described them on a 4-point scale (0 = 'Very false or often false'; 3 = 'Very true or often true'). A risk-taking score was then computed by summing the individual scores on the 14 items (range: 0–42 points). The higher the score, the higher participants' willingness to take risks. Descriptive statistics for participants' cognitive style and risk-taking measures are reported in Table 1.

2.2. Materials

Two moral dilemmas (*Surgeon* and *Bike week*; adapted from Cecchetto et al., 2017) and two exploration–exploitation problems (*Oil drilling*

Table 1. Descriptive statistics of participants' assessment measures

	Whole sample (N = 339)		NL context (N = 168)		FL context (N = 171)	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Education	16.37 (2.08)	8–24	16.15 (2.17)	8–24	16.58 (1.97)	11–22
FL AoA	5.92 (2.33)	0–19	5.97 (2.25)	0–19	5.87 (2.41)	0–14
FL normalized daily exposure	0.18 (0.17)	0–0.94	0.15 (0.15)	0–0.78	0.21 (0.18)	0–0.94
FL subjective proficiency	0.69 (0.17)	0.14–1	0.64 (0.18)	0.14–1	0.74 (0.14)	0.32–1
FL immersion	0.73 (0.1)	0.09–0.98	0.71 (0.11)	0.09–0.91	0.74 (0.08)	0.41–0.98
FL objective proficiency					17.67 (4.25)	6–25
CRT-reflective	1.73 (1.17)	0–3	1.59 (1.19)	0–3	1.88 (1.14)	0–3
Risk-taking attitude	17.56 (7.77)	1–41	17.68 (7.93)	1–41	17.44 (7.63)	2–39

Note: Means, standard deviations (*SD*) and range are reported for each measure. FL Age of Acquisition (AoA), FL Exposure normalized scores, FL subjective Proficiency and FL Immersion were collected through the Language History Questionnaire (LHQ3). The English Language Assessment (ELA) was used to collect FL objective Proficiency data from participants assigned to the FL context. The CRT-Reflective score was computed as the sum of participants' correct responses on the three Cognitive Reflection Test problems, ranging from 0 to 3. Risk-taking attitude was collected through the Italian adaptation of the Personality Inventory for DSM-5 (PID-5).

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

company and *Broken Glass*; adapted from Mehlhorn et al., 2015) were used as experimental stimuli (see, for details, Supplementary Material).

Moral dilemmas contained scenarios in which the decision-maker had to choose between killing one person to save a group of people (utilitarian option) or not killing one person and letting the group die (deontological option). The dilemmas were kept as homogeneous as possible in terms of potential moderating factors: both were *personal*, *other-beneficial*, *avoidable* and *instrumental* (see Christensen & Gomila, 2012). To rule out possible biases produced by in/out group differences (e.g., Swann et al., 2010; Uhlmann et al., 2009), the age, gender, socioeconomic status, ethnicity and cultural identity of the individuals described in each scenario were not specified. Moreover, the number of saved lives in the scenarios ranged from 5 to 7, following the categorical distinction proposed by Christensen et al. (2014) (I. 5–10; II. 11–50; III. 100–150; IV. 'thousands' or 'masses' of people). The structure (i.e., the presentation order of relevant information), the expression style (e.g., the amount of descriptive and dramatic language) and the wording of dilemmas were kept as homogeneous as possible for all scenarios and their language versions. The word count in both languages was also kept as close as possible. Outcome uncertainty was manipulated by creating two versions of each dilemma: a version formulated as deterministic in nature (outcomes of all actions were certain to happen – we refer to this version as 'certainty' condition); a version formulated as nondeterministic in nature (outcomes of all actions were uncertain to happen – we refer to this version as 'uncertainty' condition). To express different modal meanings (i.e., necessity versus possibility), two different auxiliary verbs (i.e., will/might) were assigned, respectively, to the certain and uncertain versions of each dilemma [e.g., '(…) If you transplant his organs into the bodies of the other five patients, they (will/might) be saved but the donor (will/might) die' – *Surgeon* dilemma].

Exploration–exploitation problems contained scenarios in which the decision-maker had to choose between a familiar option with a known but potentially devaluing payoff (exploitation) and an alternative option with an uncertain but potentially higher payoff (exploration). As with moral dilemmas, the problems were kept as homogeneous as possible in terms of potential moderating factors. Their structure, expression style and so forth were kept as homogeneous as possible for all scenarios and their language versions. As exploration–exploitation problems inherently involve outcome uncertainty, a single version of each problem was administered to participants.

For each language (i.e., NL, FL), two lists of problems were arranged. In the 'certainty' condition (CC) list, *Surgeon* and *Bike week* were presented in their 'certain' version, together with the *Oil drilling company* and *Broken glass* problems. In the 'uncertainty' condition (UC) list, *Surgeon* and *Bike week* were presented in their 'uncertainty' version, together with the *Oil drilling company* and *Broken glass* problems. Participants were randomly assigned to either the CC or the UC lists. In the NL context, 84 participants were assigned to the CC list and 84 to the UC list; in the FL context, 82 participants were assigned to the CC list and 89 to the UC list. Participants across the four lists did not differ in age ($\chi^2 = 3.01$, $p = .39$), gender ($\chi^2 = 5.92$, $p = .12$) or education ($\chi^2 = 5.31$, $p = .15$).

2.3. Procedure

All questionnaires were implemented via Google Forms and administered online to all participants. Debrief and instructions were given to all participants in Italian (NL). All other measurements were administered in one language, consistent with the condition participants were assigned to. Participants were asked to complete the experiment in one session. No manipulation of time pressure was exerted.

First, participants were asked to provide sociodemographic information. Then, they were presented with a problem list (CC or UC, either in the NL or FL). Afterwards, measures related to participants' FL background were collected, followed by the PID-5 and the CRT. The experimental session lasted ~40 min for participants in the NL condition, and ~ 55 min for participants in the FL condition, due to the additional measure collected for those in the FL context (i.e., the ELA). Within each list, the presentation order of the stimuli was randomized across participants. Dilemmas and problems were followed by multiple questions presented in a fixed order for all participants:

(1) A dichotomous (yes/no) question followed both moral dilemmas and exploration–exploitation problems. After reading the moral dilemmas, participants were asked whether they would choose, for each scenario, the utilitarian option of killing one to save many (*Do you decide to… yes* = utilitarian response; *no* = deontological response). After reading the exploration–exploitation problems, participants were asked whether they would choose the exploitative or the exploratory option (*Do you decide to… option 1* = exploitative response; *option 2* = exploratory response).
(2) Participants rated the extent to which dilemmas and problems made them feel distressed on three 7-point scales, each referring to a particular negative emotion or state of being (*Thinking about the scenario I just read, I felt…upset, worried, sad. 1 = not at all, 4 = somewhat, 7 = very much*).

Finally, participants assigned to the FL context were required to self-rate their understanding of each dilemma or problem presented in the FL on a 7-point scale (*Did you understand the English text in which the problem was presented?* 1 = *not at all*, 4 = *average*, 7 = *very well*). A comprehension rate ≤ 2 in at least one dilemma or problem was set as a participant's exclusion criterion.

2.4. Statistical analyses

Statistical analyses were run on R software (version 4.3.0) (R Core Team, 2015). First, a between-group analysis was used to investigate potential differences in decision-making as a function of language context and problem condition. Second, a within-group analysis was used to test whether decision-making in a FL was influenced by variability in the FL background of bilingual speakers. In addition, we tested for potential interactions between significant effects of FL background and decision-makers' traits (i.e., cognitive style and risk-taking attitude) on choice behavior.

Five participants were excluded from subsequent analyses because of poor understanding of the materials, leading to a final sample of 168 participants in the NL context and 166 in the FL context. Participants assigned to the NL and FL contexts were matched for age (W = 14723, *p-value* = .376) and gender ($\chi^2 = 2.23$, *p-value* = .135), but were slightly unbalanced in terms of education (W = 15675, *p-value* = .045). Therefore, education was entered as a covariate in the between-group analyses.

2.4.1. Between-group analysis (the effects of language context, problem condition and problem type on decision-making)

Moral dilemmas. Mixed-effect models were implemented testing the effects of language context (i.e., 'native language – NL' versus

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

'foreign language – FL') and problem condition (i.e., 'certainty – CC' versus 'uncertainty – UC'), as well as their interaction, on participants' moral decisions (utilitarian versus deontological responses) and emotional distress (distress ratings). A predictor was retained only when its inclusion determined a significant increase in explained variance. In case of a significant interaction, all the lower-order terms involved were retained. Participants and moral scenarios (i.e., *Bike week dilemma* and *Surgeon dilemma*) were modeled as random effects.

The effects of language context and problem condition on participants' moral decisions were assessed by means of a logistic mixed-effect model with type of response ('yes' versus 'no', with 'yes' codifying the utilitarian response) as a dependent variable (the logistic mixed-effect model was run using the *glmer* function in the lme4 library; Bates, Machler, Bolker, & Walker, 2015). For perceived emotional distress, since the three distress items showed good internal consistency (Cronbach's alpha was 0.85 for Bike week and 0.86 for Surgeon), we averaged them to obtain a single distress score and ran a linear mixed-effect model with the mean distress rating as a dependent variable (the linear mixed-effect model was run using the *lmer* function in the lme4 library; Bates et al., 2015). Participants' '(years of formal) education' was entered as a covariate in all statistical models.

Exploration–exploitation problems. The same analyses implemented for moral dilemmas were also implemented for exploration–exploitation problems, the only differences being the following: a) In the logistic mixed-effect models investigating participants' choices, the two levels of the dependent variable were 'exploration' and 'exploitation'; b) Since problems were presented only in one condition, there was no 'certainty' versus 'uncertainty' condition predictor; c) The two levels of the random term 'scenarios' were *Broken glass* and *Oil drilling company*. Moreover, since also in this case, the three items for emotional distress showed good internal consistency (Cronbach's alpha was 0.84 for *Oil drilling company* and 0.83 for *Broken Glass*), we averaged them to obtain a single distress score. Participants' '(years of formal) education' was entered as a covariate in all statistical models.

2.4.2. Within-group analysis (the effects of FL background on decision-making)

Before running the analyses, Spearman's correlations among measures of FL background were computed to check for any strong correlation ($r > .50$, see, e.g., Taylor, 1990) and thus avoid multicollinearity (for the correlation matrix, see Supplementary Figure S1). 'FL objective Proficiency', 'FL normalized Exposure' and 'FL Immersion' were selected as predictors. We reasoned that, as immersion takes into account active language use for extended periods of time, it represents a more valid measure of language exposure throughout the lifespan than AoA. Therefore, we entered 'FL Immersion', and not 'FL AoA', as a predictor in our models. 'FL Immersion', 'FL Exposure' and 'FL Proficiency' were centered and scaled before being entered in statistical models. We built up our models by starting from the individual and interactive effects of 'FL Proficiency', 'FL Exposure', and 'FL Immersion'. Predictors were retained only when their inclusion determined a significant increase in explained variance. When an interaction was significant, all the lower-order terms involved were retained.

Moral dilemmas. The same three dependent variables of the previous between-group analysis were investigated, i.e., the type of response ('yes' versus 'no') given to each dilemma (by means of a logistic mixed-effect model) and the emotional distress ratings (by means of a linear mixed-effect model). Participants and moral scenarios (i.e., *Bike week dilemma* and *Surgeon dilemma*) were modeled as random effects.

Exploration–exploitation problems. The same analyses implemented for moral dilemmas were also implemented for exploration–exploitation problems, the only differences being those previously described in the between-group analysis section.

2.4.3. Further analyses (the interactive effects of linguistic and psychological variables on decision-making)

For both moral dilemmas and exploration–exploitation problems, we further investigated whether significant effects of FL background were modulated by participants' cognitive style and risk-taking attitude. To do this, when a significant effect of FL background emerged, the interaction between the FL variables proved to be significant, and participants' traits were tested by entering the CRT the PID-5 scores in the final model as further predictors. Participants and scenarios were modeled as random effects. Main effects and interactions were tested as discussed above. CRT and PID-5 scores were tested in separate analyses. Both the CRT and the PID-5 scores were centered and scaled before being entered into statistical models.

3. Results

3.1. Between-group analysis (the effects of language context and problem condition on decision-making)

3.1.1. Moral dilemmas

For each moral dilemma, the proportion of utilitarian responses and the mean emotional distress ratings as a function of language context and problem condition are represented in Figure 1. Significant effects of tested predictors are reported separately for each dependent variable.

Moral decisions (binary responses). No effect of language context ($\chi^2 = .504$, *p-value* = .478, beta = −.155, st.err. = .215, z-value = .723) or problem condition ($\chi^2 = 2.7$, *p-value* = .1, beta = −.36, st.err. = .216, z-value = −1.666), nor their interaction ($\chi^2 = 3.387$, *p-value* = .066, beta = −.806, st.err. = .433, z-value = −1.860), on participants' moral decisions reached significance.

Emotional distress (rating scales). The main effect of language context ($\chi^2 = 4.736$, *p-value* = .030) on participants' distress ratings reached significance. Specifically, a lower emotional distress was associated with NL (versus FL) processing (beta = −0.380, st. err. = 0.174, t-value = −2.179). Neither the effect of problem condition ($\chi^2 = 2.603$, *p-value* = .107, beta = −.280, st.err. = .174, t-value = −2.213) nor the interaction ($\chi^2 = .304$ *p-value* = .581, beta = −.191, st.err. = .348, t-value = −.549) reached significance.

3.1.2. Explorations versus exploitation problems

For each exploration–exploitation problem, the proportion of explorative responses and the mean emotional distress ratings as a function of language context are represented in Figure 1. Significant effects of tested predictors are reported separately for each dependent variable.

Explorations versus exploitation choices (binary responses). No significant effect of language context on participants' decisions was observed ($\chi^2 = 1.804$, *p-value* = .191, beta = −.293, st.err. = .216, z-value = −1.356).

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

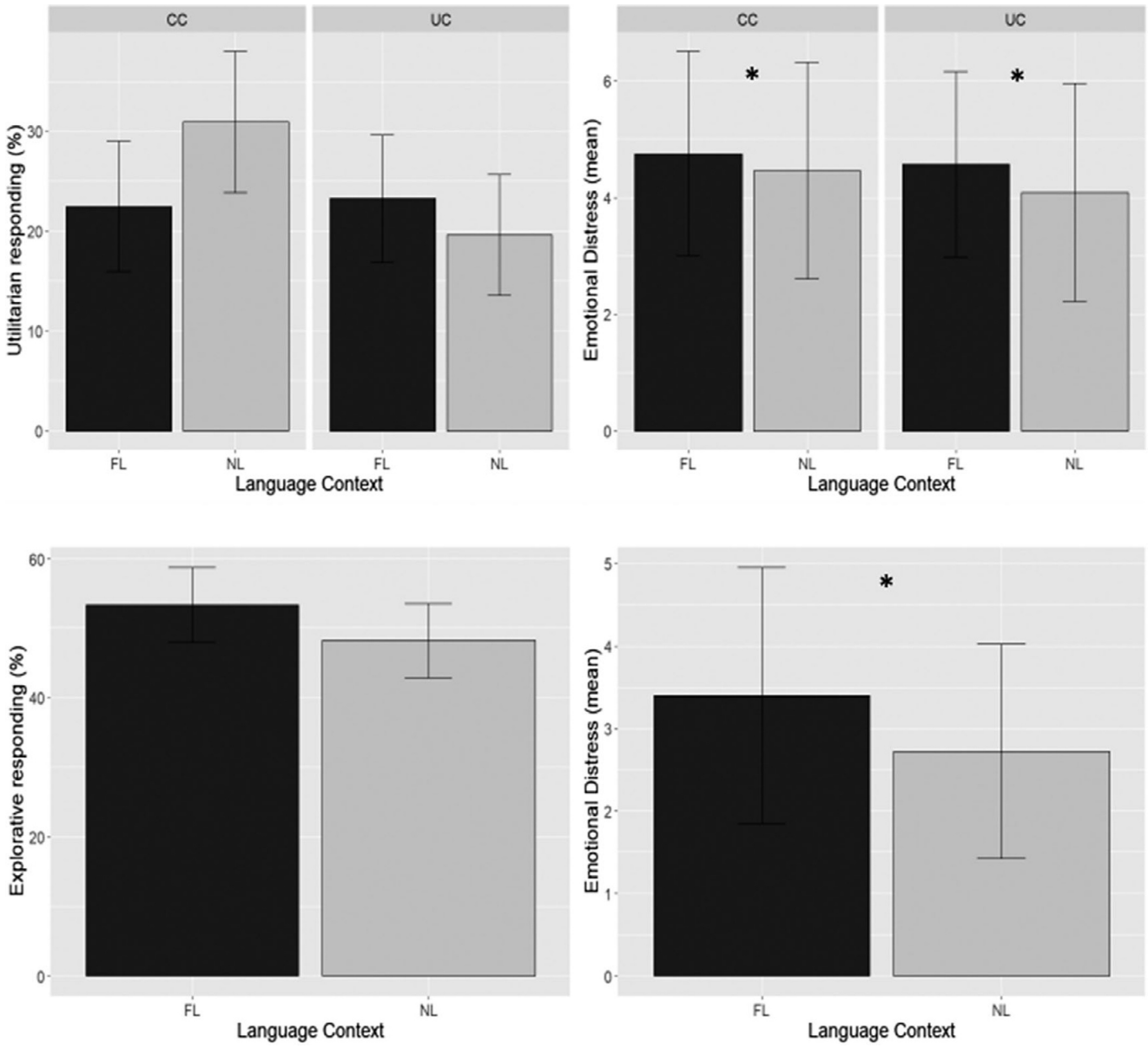


Figure 1. Effects of language context and problem condition on participants' decision-making and perceived emotional distress. For moral dilemmas (on top), the percentage of utilitarian responses (left panel) and the mean emotional distress ratings (right panel) are reported as a function of language context (FL = foreign language; NL = native language) and problem condition (CC = certainty condition; UC = uncertainty condition). For exploration–exploitation problems, the percentage of explorative responses (left panel) and the mean emotional distress ratings (right panel) are reported as a function of language context. Error bars represent the standard error of the mean. Significant differences between language contexts are marked with an asterisk ($p < .05$).

Emotional distress (rating scales). A significant effect of language context (χ^2 = 21.078, *p-value* < .001) on participants' emotional distress was detected, with lower ratings associated with NL (versus FL) processing (beta = −0.673, st. err. = 0.145, t-value = −4.651)[1].

3.2. Within-group analysis (the effects of FL background on decision-making)

3.2.1. Moral dilemmas

For each moral dilemma, significant effects of FL background are reported separately for each dependent variable.

Moral decisions (binary responses). A significant three-way interaction between FL proficiency, FL exposure and FL immersion on participants' moral decisions was observed ($\chi^2(1)$ = 5.719, *p-value* = .017, beta = −.379, st. err. = .216, z-value = −1.752; see Supplementary Figure S2 and Supplementary Table S1). In particular, for participants with a low immersive experience with the FL, the probability of choosing the utilitarian (versus deontological) option increased with increasing proficiency when FL exposure at the time of testing was high, whereas the same probability decreased when FL exposure was low. For participants with a highly immersive experience with the FL, the probability of choosing the utilitarian (versus deontological) option decreased with increasing proficiency when FL exposure at the time of testing was high, while the same probability increased when FL exposure was low.

[1]As the estimated linear mixed-effect model was singular, the random term explaining the lowest variance, i.e., *scenarios*, was removed from the fitted model.

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

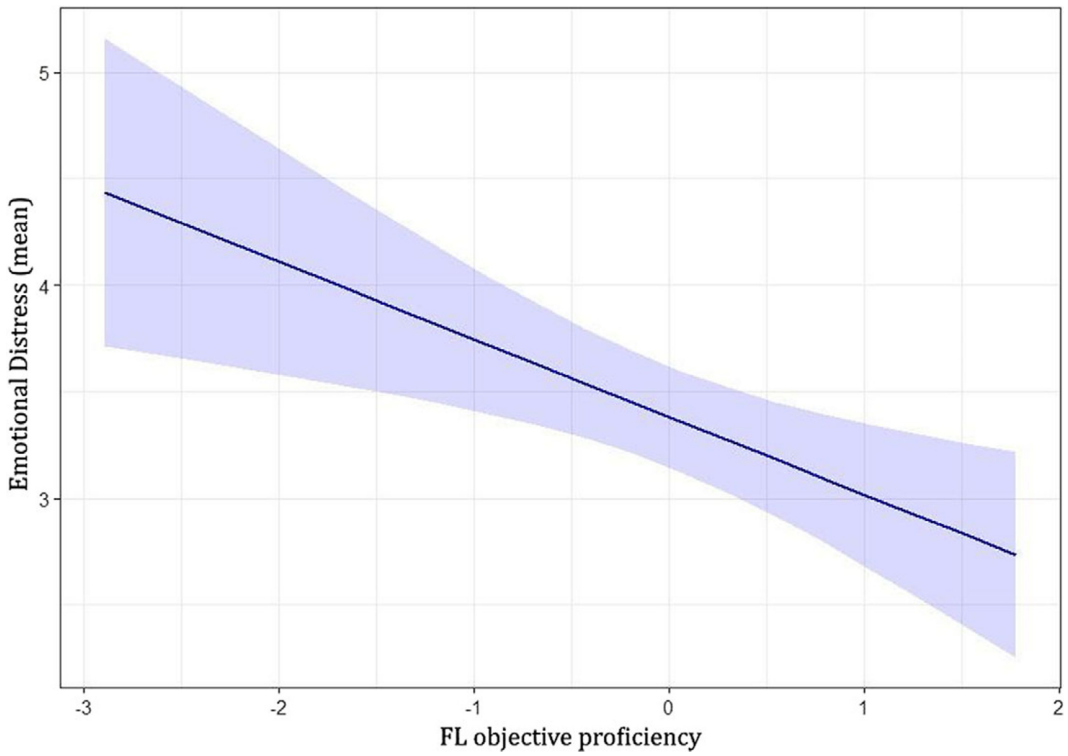


Figure 2. Effects of foreign language (FL) proficiency on participants' perceived emotional distress in exploration—exploitation problems. The figure represents the main effect of objective proficiency in the FL (ELA score) on the emotional distress (mean emotional distress ratings) associated with processing exploration–exploitation problems. Values represent the predicted values conditioned on FL proficiency derived from the fitted model. The FL proficiency predictor is represented as scaled and centered.

Emotional distress (rating scales). The three-way interaction between FL proficiency, FL exposure and FL immersion on participants' emotional distress ratings reached significance ($\chi^2(1) = 4.892$, *p-value* = .027, beta = −.197, st. err. = .091, t-value = −2.171; see Supplementary Figure S3 and Supplementary Table S2). In particular, for participants with a low immersive experience with the FL, the perceived emotional distress increased with increasing proficiency when FL exposure at the time of testing was high, whereas it decreased when FL exposure was low. The reverse pattern was observed for participants with a highly immersive experience with the FL.

3.2.2. Exploration–exploitation problems

Exploration versus exploitation choices (binary responses). A significant two-way interaction between FL proficiency and FL exposure on participants' explorative (versus exploitative) responses was observed ($\chi^2(1) = 4.867$, *p-value* = .027, beta = .366, st.err. = .175, z-value = 2.096; see Supplementary Table S3). In particular, for participants with a high exposure to the FL at the time of testing, the probability of choosing the explorative (versus exploitative) option increased with increasing proficiency, whereas the same probability decreased when FL exposure was lower. The FL proficiency x FL immersion interaction on participants' choices also reached significance ($\chi^2(1) = 4.999$, *p-value* = .025, beta = .419, st.err. = .194, z-value = 2.161; see Supplementary Figure S4 and Supplementary Table S3). The interaction showed that the probability of explorative responses increased with increasing proficiency for participants with a highly immersive experience with the FL, whereas it decreased when the degree of FL immersion was lower. No other effect reached significance (all *p-values* > .40).

Emotional distress (rating scales). A main effect of FL proficiency on participants' emotional distress ratings emerged ($\chi^2(1) = 9.031$, *p-value* = .003), revealing that participants with a higher proficiency level experienced a lower emotional distress (beta = −.364, st.err. = .120, t-value = −3.033; see Figure 2)[2]. No other effect reached significance (all *p-values* > .09; Supplementary Table S4).

3.2.3. Further analyses (the interactive effects of linguistic and psychological variables on decision-making)

The effects of FL background that emerged as significant in previous analyses were further inspected in interaction with cognitive style (CRT) and risk-taking (PID-5) measures.

3.2.4. Reflective thinking (CRT score)

Moral dilemmas: Moral decisions (binary responses). The CRT score was added into the final model, including the three-way interaction between FL proficiency, FL exposure and FL immersion. The FL proficiency x FL exposure x FL immersion interaction on participants' moral responses was still significant ($\chi^2(1) = 10.808$, *p-value* = .001, beta = −.525, st.err. = .210, z-value = −2.503). Furthermore, a significant FL proficiency x FL immersion x CRT score interaction was observed ($\chi^2(1) = 8.815$, *p-value* = .003, beta = −.739, st. err. = .272, z-value = −2.720; see Supplementary Figure S5 and Supplementary Table S5). For participants with a low propensity toward reflective (versus intuitive) thinking, the probability of choosing the utilitarian (versus deontological) option increased with increasing proficiency when experience with the FL was highly immersive, while the same probability decreased when the degree of FL immersion was lower. In participants with a high propensity toward reflective (versus intuitive) thinking, the probability of

[2]See footnote 1.

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

choosing the utilitarian (versus deontological) option decreased with increasing proficiency when experience with the FL was highly immersive, whereas it increased when the degree of FL immersion was lower.

Moral dilemmas: Emotional distress (rating scales). The CRT score was added into the final model, including the three-way interaction between FL proficiency, FL exposure and FL immersion. The FL proficiency x FL exposure x FL immersion interaction on participants' moral responses was still significant ($\chi^2(1) = 6.776$, *p-value* = .009, beta = −.233, st.err. = .092, t-value = −2.545). Furthermore, a significant FL immersion x CRT score interaction was observed ($\chi^2(1) = 4.123$, *p-value* = .042; see Supplementary Figure S6 and Supplementary Table S6), indicating that the perceived emotional distress decreased with increasing FL immersion for participants with a high propensity toward reflective (versus intuitive) thinking, whereas it increased when the propensity toward reflective (versus intuitive) thinking with lower (beta = −.271, st.err. = .137, t-value = −1.976).

Exploration–exploitation problems: Exploration versus exploitation choices (binary responses). When the CRT score was added into the final model, both the two-way interaction between FL proficiency and FL exposure ($\chi^2(1) = 4.869$, *p-value* = .027, beta = .366, st.err. = .175, z-value = 2.096) and the two-way interaction between FL proficiency and FL immersion ($\chi^2(1) = 4.999$, *p-value* = .025, beta = .419, st.err. = .194, z-value = 2.161) were still significant. No other effect reached significance (all $\chi^2 < 1$, all *p-values* > .20).

Exploration–exploitation problems: Emotional distress (rating scales). When the CRT score was added into the final model, the main effect of FL proficiency on participants' emotional distress ratings was still significant ($\chi^2(1) = 9.031$, *p-value* = .003, beta = −.364, st.err. = .120, t-value = −3.033). No other significant effect was observed (all $\chi^2 < 1.5$, all *p-values* > .20).

3.2.5. Risk-taking attitude (PID-5)

Moral dilemmas: Moral decisions (binary responses). The PID-5 score was added as a further predictor in the final model, including the three-way interaction between FL proficiency, FL exposure and FL immersion. The FL proficiency x FL exposure x FL immersion interaction was still significant ($\chi^2(1) = 4.311$, *p-value* = .038, beta = −.414, st.err. = .331, z-value = −1.251). Moreover, a two-way interaction between FL proficiency and risk-taking attitude ($\chi^2(1) = 9.157$, *p-value* = .002), showing that the probability of choosing the utilitarian (versus deontological) option decreased with increasing proficiency for participants with a higher risk-taking attitude, while the same probability increased for participants with a lower risk-taking attitude (beta = −.520, st.err. = .184, t-value = −2.817; see Figure 3 and Supplementary Table S7).

Moral dilemmas: Emotional distress (rating scales). The PID-5 score was added as a further predictor in the final model, including the three-way interaction between FL proficiency, FL exposure and FL immersion. The three-way interaction between FL proficiency, FL exposure and FL immersion was no more significant ($\chi^2(1) = 2.184$, *p-value* = .139). Nevertheless, a significant three-way interaction between FL proficiency, FL exposure and risk-taking attitude was observed ($\chi^2(1) = 7.498$, *p-value* = .006, beta = −.391, st.err. = .145, z-value = −2.701; see Supplementary Figure S7 and

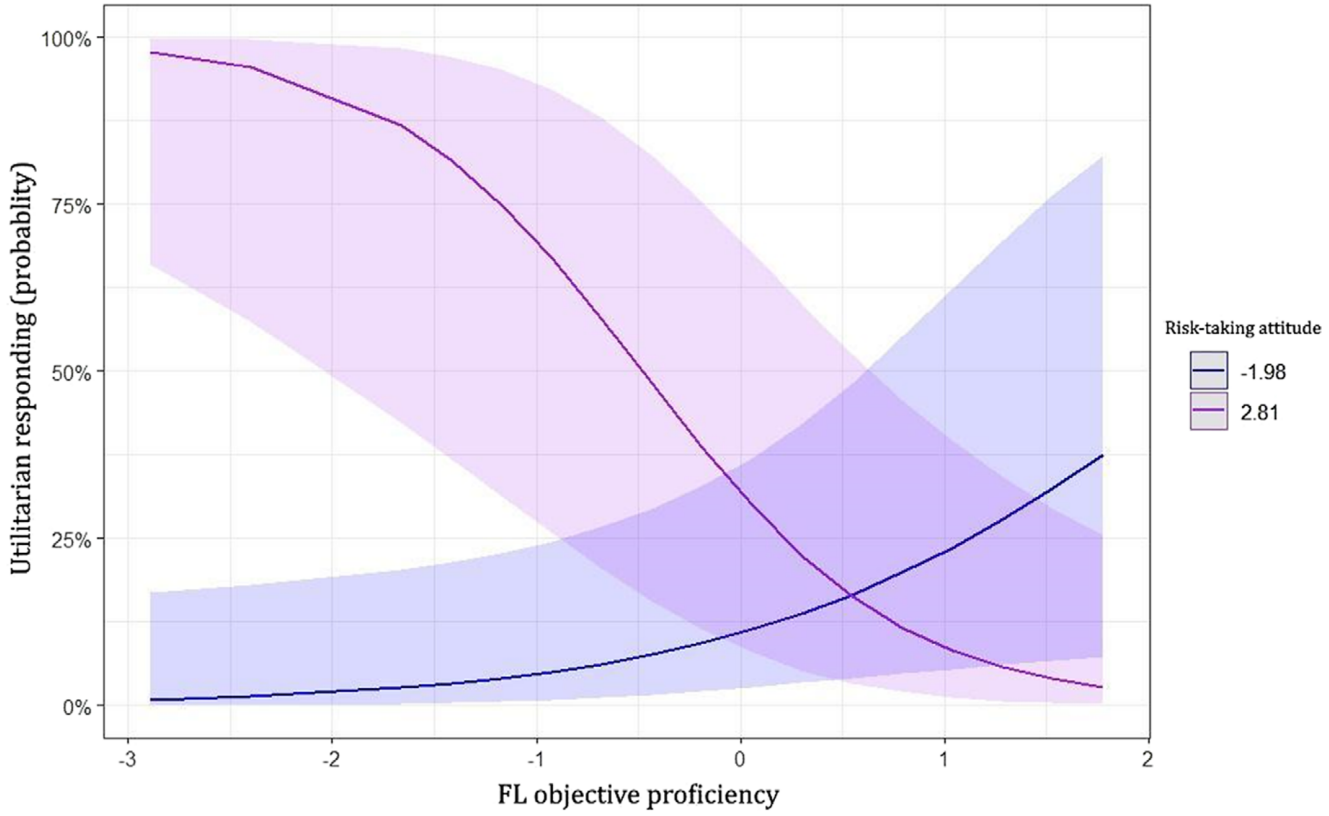


Figure 3. Interactive effects of foreign language (FL) proficiency and risk-taking attitude on participants' moral choices.The figure represents the two-way interaction between objective proficiency in the FL (ELA score) and risk-taking attitude (PID-5 score) on participants' moral choices (i.e., number of utilitarian responses in moral dilemmas). Values represent the predicted probability conditioned on the fixed effect terms (i.e., FL proficiency and risk-taking attitude) specified in the fitted model. Minimum and maximum values (i.e., lower and upper bounds) of the PID-5 score were used to plot the interaction. Both predictors are represented as scaled and centered.

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

Supplementary Table S8). In particular, in participants with a lower risk-taking attitude, the perceived emotional distress marginally increased with increasing proficiency when FL exposure at the time of testing was high, whereas it marginally decreased when FL exposure was lower. In participants with a higher risk-taking attitude, emotional distress decreased with increasing proficiency when FL exposure was high, while it marginally increased when FL exposure was lower.

Exploration–exploitation problems: Exploration versus exploitation choices (binary responses). The PID-5 score was added as a further predictor in the final model including a two-way interaction between FL proficiency and FL exposure and a two-way interaction between FL proficiency and FL immersion. Both the FL proficiency × FL exposure interaction ($\chi^2(1) = 4.175$, *p-value* = .041, beta = .337, st.err. = .173, z-value = 1.946) and the FL proficiency x FL immersion interaction ($\chi^2(1) = 6.012$, *p-value* = .014, beta = .461, st.err. = .197, z-value = 2.347) were still significant. Furthermore, the main effect of risk-taking attitude reached significance ($\chi^2(1) = 4.399$, *p-value* = .036; see Supplementary Table S9), with a greater tendency to choose the explorative (versus exploitative) choice in participants with a higher risk-taking attitude (beta = .340, st.err. = .165, z-value = 2.058).

Exploration–exploitation problems: Emotional distress (rating scales). The PID-5 score was added as a further predictor into the final model, including FL proficiency. The main effect of FL proficiency on participants' emotional distress ratings was still significant ($\chi^2(1) = 9.031$, *p-value* = .003, beta = −.364, st.err. = .120, t-value = −3.033). No other effect reached significance (all $\chi^2 < 3$, all *p-values* > .09).

4. Discussion

Uncertainty is ubiquitous in realistic settings, yet research on the FLE has largely overlooked whether the lack of information about the probability of future events affects people's choices in a FL. In the next paragraphs, we first discuss participants' differences in choice behavior as a function of language context and problem condition. Then, we discuss the role of emotional distress on participants' decisions in light of emotion-based accounts of the FLE. We then examine whether decision-making in a FL was influenced by variability in the language experiences of our bilingual speakers. Finally, we discuss significant interactions between linguistic and psychological factors on participants' decision-making. We conclude by outlining the relevance of our results for research on the FLE.

4.1. Between-group analysis (the effects of language context and problem condition on decision-making)

Overall, our between-groups analyses showed null effects of language context (NL versus FL) and problem condition (certainty versus uncertainty) on participants' decision-making (for both moral dilemmas and exploration–exploitation problems).

At present, the available evidence regarding the putative effects of language context on moral decision-making is mixed. While especially early investigations reported a greater preference for utilitarian over deontological options when making judgments in a FL, such evidence did not consistently replicate in more recent studies (e.g., Białek, Paruzel-Czachura, & Gawronski, 2019; Del Maschio, Del Mauro, et al., 2022b; Feng et al., 2023; Hayakawa et al., 2017; Muda, Walker, et al., 2020; Nadarevic et al., 2021). In addition, contrary to our predictions, response patterns to moral dilemmas were very similar under certain and uncertain prospects in both languages. Problem condition was manipulated by creating, for each dilemma, a 'deterministic' version (the auxiliary verb 'will' was used to indicate that the outcomes of all prospected actions were certain to happen) and a nondeterministic version (the auxiliary verb 'might' was used to indicate that the outcomes of all prospected actions were uncertain). In interpreting our null finding of problem condition on moral decisions, we cannot rule out the possibility that our linguistic manipulation was too small to influence the decision-making processes of our participants.

It has been suggested that risks appear smaller in a FL. For instance, potential hazards associated with various activities and technologies (e.g., traveling by airplane, chemical fertilizers) are perceived as less risky and more beneficial when presented in a FL compared to a NL (Hadjichristidis et al., 2015). On these grounds, we predicted that, when confronted with a trade-off between repeating a past action in expectation of a familiar outcome (exploitation) and a novel action whose outcome is uncertain but potentially of superior value (exploration), participants assigned to a FL context would be more willing to engage in riskier, exploratory behavior than those assigned to the NL context. However, although the use of a FL prompted a marginally higher proportion of explorative (versus conservative) decisions than the NL across exploration–exploitation problems, such differences failed to reach statistical significance.

4.2. A reduced emotional resonance when making decisions in a FL?

The FLE has been mostly interpreted as emerging from a reduction in emotional processing. In particular, especially in unbalanced bilinguals, operating in a FL is supposed to elicit higher emotional distancing compared to the NL (see Caldwell-Harris, 2015). In decision-making contexts, a weaker emotional resonance of the FL is expected to reduce the likelihood of affect-based responses to decision problems in favor of more deliberative cost–benefit appraisals (e.g., Cipolletti et al., 2016; Costa, Foucart, Hayakawa, et al., 2014b). Our results do not support this hypothesis. Indeed, our participants reported perceiving a higher emotional distress in the FL than in the NL, in sharp contrast with the emotional distancing hypothesis. Of note, this pattern was relatively consistent across dilemmas and problems, and independent of problem condition. This result is unexpected and adds to other problematic findings that defy predictions based on the emotion-reducing effect of FLs in decision-making contexts (e.g., Chan et al., 2016; Del Maschio, Del Mauro, et al., 2022b; Geipel et al., 2015; Miozzo et al., 2020; Muda, Walker, et al., 2020). Chan et al. (2016), for example, administered a large number of moral dilemmas to a group of Chinese–English bilinguals and tested whether emotional arousal played a mediating role in the effect of language on moral choices. The authors failed, in the first place, to report any significant relationship between language and moral choices except for the Footbridge dilemma. Crucially, this stimulus-specific effect of language was not mediated by emotional arousal, in contrast with the emotional distancing hypothesis. Similar findings were previously obtained by Geipel et al. (2015), who found an effect of language on moral judgment but failed to observe a mediating role of emotion. It is noteworthy that our findings do not simply testify a lack of connection between FLE and emotional processing, but suggest that processing decision problems in a FL generates higher emotional distress as compared to NL contexts. A tentative explanation for this result relates perceived emotional distress to cognitive effort. The mere fact of making a high-conflict decision in a FL,

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press

in a sample of participants who speak a FL with varying degrees of proficiency and exposure, but who were born and raised in environments in which their NL was dominantly spoken, may increase the cognitive effort associated with processing a FL text, making processing more frustrating or distressing. However, at least two elements seem to weaken this interpretation. First, the increased cognitive effort associated with processing problems in a FL is expected to lead to greater reliance on intuitive and affective processes, exacerbating decision biases (e.g., Costa et al., 2014a; Hayakawa et al., 2017; cf. also the 'reduced-systematicity account' in Keysar et al., 2012). This is not warranted by the lack of significant differences between participants' choices as a function of language context. Second, cognitive effort is unlikely to be associated with feelings or moods like being 'upset', 'sad' or 'worried' (i.e., the probes of our emotional distress questions). We speculate that other kinds of feelings or moods (e.g., 'frustration') are more likely to be associated with cognitive effort or overload. Future research may investigate the mechanisms underlying the putative interplay between cognitive effort, FL context and emotional processing by designing questions specifically aimed at assessing the emotional correlates of cognitive effort in FL versus NL contexts. On the whole, in light of the null effects of language context and problem condition on bilinguals' decision-making, the only inference that seems safe to draw from this particular set of findings is that the emotional distancing hypothesis is not suitable to explain the higher distress in the FL (versus NL) that we consistently report across problems.

4.3. Within-group analyses (the effects of FL background and psychological variables on decision-making)

We operationalized bilingualism as a construct comprising several interrelated dimensions and adopted a perspective that takes into account the extent to which individuals vary as bilinguals along FL proficiency, FL exposure and FL immersion. As predicted, these variables were found to interact in modulating decision outcomes across both moral dilemmas and exploration–exploitation problems. However, the modulatory role of specific components of bilinguals' language experience emerged for some problems but not others, and in the face of a null effect of language context on those same problems. This pattern of findings suggests that differences in the language experiences of bilingual speakers can influence bilinguals' choices in a FL, without necessarily emerging into the 'classic' FLE on decision-making. Importantly, when effects of FL background were detected, the direction of such effects on decision outcomes and distress ratings was not always consistent across problems, and in some cases, incompatible with previous theorizing. For example, in line with the hypothesis that high levels of proficiency in the FL would promote emotional grounding (see Hayakawa et al., 2016), we predicted that increased proficiency would also be associated with higher emotional distress when processing problems in one's FL. Contrary to our expectations, but consistent with the pattern of findings obtained from our between-group analysis, when processing exploration–exploitation problems, participants with higher FL proficiency reported to experience a lower emotional distress than participants with a higher proficiency level.

When we tested for potential interactions between significant effects of FL background and decision-makers' traits on choice behavior, we found that effects of variables related to bilinguals' language experience were influenced by individuals' cognitive style (CRT scores) and dispositions toward risk (PID-5 scores). Importantly, these findings suggest that relatively stable aspects of cognitive processing and personality can interact with more dynamic variables related to bilingual language experience to shape individuals' choices.

There are a number of limitations to the present study. A potential limitation is that we used, as experimental stimuli, scenarios that may be unrepresentative of situations people can face on an ordinary basis. Our moral dilemmas, in particular, are sacrificial dilemmas that have been criticized for low likelihood of occurrence (e.g., Bauman et al., 2014). On the one hand, however, we introduced uncertainty in our decision-making contexts with the specific purpose of increasing the mundane and psychological realism of our stimulus materials. On the other hand, the use of simple artificial settings allowed us to exert a higher experimental control on conceptual and methodological aspects of problem design that have been shown to influence people's judgment (Christensen & Gomila, 2012). Furthermore, given the online nature of the present study and the subtlety of the manipulations involved, future research may use more advanced online survey tools to monitor participants' behavior while completing the experiment, thereby enhancing data reliability. Another limitation to the present study, which is inherent in our between-group analysis, is that the absence of observed group differences may reflect overlapping variances rather than true null effects. Finally, although our participants' sample is larger than the samples used in previous research on the FLE with an approach similar to the one adopted here (Kirova et al., 2023; Privitera, 2024; Privitera et al., 2023), a larger sample size would provide a stronger test for the conclusions suggested by the present results.

5. Conclusions

The FLE has been mostly tested in contexts where uncertainty is expunged or reduced to a form of risk. We explored whether the FLE on decision-making extends to uncertain scenarios in order to provide a more ecological picture of the FLE on choice behavior. Overall, we failed to detect any effect of language context (NL versus FL) or problem condition (certainty versus uncertainty) on participants' decision-making. In addition, we found that both FL background and decision makers' traits modulated participants' choices in a FL, without emerging into the 'classic' FLE on decision-making. However, the direction of such effects was complex and not always compatible with previous FLE theories. This overall pattern of findings calls for a general rethinking of the phenomenon and its underlying mechanisms.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925100400.

Data availability statement. The raw data supporting the findings of this study are available at https://osf.io/2m9nk/?view_only=0ea16144427742ebb8b58e4d8369a95c. Scripts used to analyze the data are available upon request to the corresponding author.

Competing interests. The authors declare none.

References

Bauman, C. W., **McGraw, A. P.**, **Bartels, D. M.**, & **Warren, C.** (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, **8**, 536–554.

Bates, D., **Machler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.

Białek, M., **Paruzel-Czachura, M.**, & **Gawronski, B.** (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, **85**, 103855.

Caldwell-Harris, C. L. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science*, **24**, 214–219.

Cecchetto, C., **Rumiati, R. I.**, & **Parma, V.** (2017). Promoting cross-culture research on moral decision making with standardized, culturally-equivalent dilemmas: The 4CONFiDe set. *Journal of Health and Social Sciences*, **22**, 173–194.

Chan, Y. L., **Gu, X.**, **Ng, J. C. K.**, & **Tse, C. S.** (2016). Effects of dilemma type, language, and emotion arousal on utilitarian vs deontological choice to moral dilemmas in Chinese–English bilinguals. *Asian Journal of Social Psychology*, **19**, 55–65.

Christensen, J. F., & **Gomila, A.** (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience & Biobehavioral Reviews*, **36**, 1249–1264.

Christensen, J. F., **Flexas, A.**, **Calabrese, M.**, **Gut, N. K.**, & **Gomila, A.** (2014). Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology*, **5**, 607.

Cipolletti, H., **McFarlane, S.**, & **Weissglass, C.** (2016). The moral foreign-language effect. *Philosophical Psychology*, **29**, 23–40.

Costa, A., **Foucart, A.**, **Arnon, I.**, **Aparici, M.**, & **Apesteguia, J.** (2014a). Piensa" twice: On the foreign language effect in decision making. *Cognition*, **130**, 236–254.

Costa, A., **Foucart, A.**, **Hayakawa, S.**, **Aparici, M.**, **Apesteguia, J.**, **Heafner, J.**, & **Keysar, B.** (2014b). Your morals depend on language. *PLoS One*, **9**, e94842.

Costa, A., **Vives, M. L.**, & **Corey, J. D.** (2017). On language processing shaping decision making. *Current Directions in Psychological Science*, **26**, 146–151.

Del Maschio, N., **Crespi, F.**, **Peressotti, F.**, **Abutalebi, J.**, & **Sulpizio, S.** (2022a). Decision-making depends on language: A meta-analysis of the foreign language effect. *Bilingualism: Language and Cognition*, **25**, 617–630.

Del Maschio, N., **Del Mauro, G.**, **Bellini, C.**, **Abutalebi, J.**, & **Sulpizio, S.** (2022b). Foreign to whom? Constraining the moral foreign language effect on bilinguals' language experience. *Language and Cognition*, **14**, 511–533.

Díaz-Lago, M., & **Matute, H.** (2019). Thinking in a foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, **72**(1), 41–51.

Erceg, N., & **Bubić, A.** (2017). One test, five scoring procedures: Different ways of approaching the cognitive reflection test. *Journal of Cognitive Psychology*, **29**(3), 381–392.

Feng, C., **Liu, C.**, & **Nolasco, E. E.** (2023). Revisiting the foreign language effect on moral decision-making: A closer look at the role of perverse responses. *Journal of Multilingual and Multicultural Development*, 1–19.

Fossati, A., & **Somma, A.** (2021). The assessment of personality pathology in adolescence from the perspective of the alternative DSM-5 Model for Personality Disorder. *Current Opinion in Psychology*, **37**, 39–43.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, **19**(4), 25–42.

Geipel, J., **Hadjichristidis, C.**, & **Surian, L.** (2015). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS One*, **10**, e0131529.

Gilovich, T., **Griffin, D.**, & **Kahneman, D.** (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

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**, 117.

Hadjichristidis, C., **Geipel, J.**, & **Surian, L.** (2019). Breaking magic: Foreign language suppresses superstition. *Quarterly Journal of Experimental Psychology*, **72**, 18–28.

Hayakawa, S., **Costa, A.**, **Foucart, A.**, & **Keysar, B.** (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*, **20**, 791–793.

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**, 1387–1397.

Keller, K. L., & **Staelin, R.** (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research*, **14**(2), 200–213.

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

Kirova, A., **Tang, Y.**, & **Conway, P.** (2023). Are people really less moral in their foreign language? Proficiency and comprehension matter for the moral foreign language effect in Russian speakers. *PLoS One*, **18**(7), e0287789.

Kortenkamp, K. V., & **Moore, C. F.** (2014). Ethics under uncertainty: The morality and appropriateness of utilitarianism when outcomes are uncertain. *The American Journal of Psychology*, **127**, 367–382.

Krueger, R. F., **Derringer, J.**, **Markon, K. E.**, **Watson, D.**, & **Skodol, A. E.** (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, **42**, 1879–1890.

Li, P., **Zhang, F.**, **Yu, A.**, & **Zhao, X.** (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, **23**(5), 938–944.

Mækelæ, M., & **Pfuhl, G.** (2019). Deliberate reasoning is not affected by language. *PLoS One*, **14**, e0211428.

Mehlhorn, K., **Newell, B. R.**, **Todd, P. M.**, **Lee, M. D.**, **Morgan, K.**, **Braithwaite, V. A.**, & **Gonzalez, C.** (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, **2**(3), 191.

Miozzo, M., **Navarrete, E.**, **Ongis, M.**, **Mello, E.**, **Girotto, V.**, & **Peressotti, F.** (2020). Foreign language effect in decision-making: How foreign is it? *Cognition*, **199**, 104245.

Mousavi, S., & **Gigerenzer, G.** (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, **67**, 1671–1678.

Muda, R., **Walker, A. C.**, **Pieńkosz, D.**, **Fugelsang, J. A.**, & **Białek, M.** (2020). Foreign language does not affect gambling-related judgments. *Journal of Gambling Studies*, **36**, 633–652.

Nadarevic, L., **Klein, L. C.**, & **Dierolf, J.** (2021). Does foreign language alter moral judgments? Inconsistent results from two pre-registered studies with the CNI model. *Open Psychology*, **3**, 66–86.

Paxton, J. M., **Ungar, L.**, & **Greene, J. D.** (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, **36**, 163–177.

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.

Privitera, A. J. (2024). Is there a foreign language effect on academic integrity? *Higher Education*, **88**(2), 609–626.

Purpuri, S., **Vasta, N.**, **Filippi, R.**, **Wei, L.**, & **Mulatti, C.** (2024a). Does language shape the way we think? A review of the foreign language effect across domains. *International Journal of Bilingualism*, 13670069231225374.

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.

R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Sulpizio, S., **Del Maschio, N.**, **Del Mauro, G.**, **Fedeli, D.**, & **Abutalebi, J.** (2020). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *NeuroImage*, **205**, 116306.

Sulpizio, S., **Toti, M.**, **DelMaschio, N.**, **Costa, A.**, **Fedeli, D.**, **Job, R.**, & **Abutalebi, J.** (2019). Are you really cursing? Neural processing of taboo words in native and foreign language. *Brain and Language*, **194**, 84–92.

Swann, W. B., **Gomez, A.**, **Dovidio, J. F.**, **Hart, S.**, & **Jetten**. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science*, **21**,1176–1183.

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, **6**(1), 35–39.

Uhlmann, E. L., **Pizarro, D. A.**, **Tannenbaum, D.**, & **Ditto, P. H.** (2009). The motivated use of moral principles. *Judgment and Decision Making*, **4**, 476–491.

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**, e0203528.

Wittmann, B. C., **Daw, N. D.**, **Seymour, B.**, & **Dolan, R. J.** (2008). Striatal activity underlies novelty-based choice in humans. *Neuron*, **58**, 967–973.

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.

Yang, X., **Li, L.**, & **Li, R.** (2021). Foreign language effect on dishonesty. *Frontiers in Psychology*, **12**, 633016.

https://doi.org/10.1017/S1366728925100400 Published online by Cambridge University Press