



City Research Online

City St George's, University of London

Citation: Dylman, A. & Samuel, S. (2026). Does linguistic similarity matter in the foreign language effect in decision making and emotional resonance?. *Language and Cognition*, 18, e27. doi: 10.1017/langcog.2026.10077

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/37070/>

Link to published version: <https://doi.org/10.1017/langcog.2026.10077>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Does linguistic similarity matter in the foreign language effect in decision making and emotional resonance?

Alexandra S. Dylman¹

Steven Samuel²

¹School of Education, Culture and Communication, Mälardalen University, Sweden

²City St. George's, Department of Psychology and Neuroscience, School of Health and Medical Sciences, University of London, UK

Author Note

Correspondence concerning this article should be addressed to Alexandra Dylman, School of Education, Culture and Communication, Mälardalen University, Universitetsplan 1, 722 20 Västerås, Sweden. Email: asdylman@gmail.com

Abstract

This study investigated the role of linguistic similarity in the foreign language effect (FLE) on decision-making and emotional resonance. Previous research has found that using a foreign language (L2) leads to reduced emotionality and increased rationality compared to a first language (L1). We report two experiments with different target languages: Experiment 1 investigated L1 and L2 speakers of English, and Experiment 2 investigated L1 and L2 speakers of Swedish. Participants in both experiments completed three decision-making tasks, including the Asian disease problem, a moral dilemma, and the Cognitive reflection test. The L2 speakers also reported their emotional resonance using the Reduced Emotional Resonance in LX scale (RER-LX). Results from Experiment 1 showed no significant FLE in the decision-making tasks but indicated that linguistic similarity affects emotional resonance. Experiment 2 found a classic FLE in all decision-making tasks, with participants in L2 contexts making more rational and utilitarian decisions. However, linguistic similarity did not consistently affect the magnitude of the FLE. The study highlights the complexity of the FLE and suggests that factors such as age of acquisition and immersion may influence its manifestation.

Keywords: bilingualism, emotional resonance, foreign language effect, linguistic similarity, second language use

Does linguistic similarity matter in the foreign language effect in decision making and emotional resonance?

Over the past decade or so, how language context affects people's behaviours, judgments and decision making, and emotions, has been a prevalent subject of investigation in research on multilingual cognition. Specifically, studies have found that a first language (L1) context is associated with larger emotional resonance and reactivity, while a second- or foreign language context (L2) is associated with attenuated emotionality (e.g., Caldwell-Harris, 2014; Dewaele, 2010; Pavlenko, 2006). Relatedly, people make more rational decisions and are less likely to fall for decision making biases in an L2 compared to an L1 context (see e.g., Circi et al., 2021 for an overview). For example, while people are typically affected by how problems are worded (e.g., as in the well-known Asian disease problem), this framing effect is diminished or even eliminated in an L2 context (e.g., Keysar et al., 2012). Similar effects of language context on various other domains of decision making have also been found, including observations that people are less risk seeking and more risk averse in an L1 compared to an L2, or show a lower tendency to fall for heuristic biases in an L2 compared to an L1 (e.g., Costa et al., 2014a, Hayakawa et al., 2019; Keysar et al., 2012).

Additionally, research has found an effect of language context on people's responses when faced with hypothetical moral dilemmas (e.g., Corey et al., 2017; Costa et al., 2014b). When asked to make decisions about scenarios about sacrificing one life to save those of many, the majority of people tend to choose the deontological option over the utilitarian option of maximising the global benefit. In other words, they decide against actively killing one person, even if it would mean saving multiple other lives, instead adhering to deep-rooted moral codes (such as "do not kill" in this case). However, there is typically a small proportion of people in the population who will choose the utilitarian option when faced with these types

of hypothetical moral dilemmas. Interestingly, this proportion tends to be significantly larger in an L2 context (e.g., Corey et al., 2017; Costa et al., 2014b).

This phenomenon of reduced emotionality and increased rationality in an L2 is commonly referred to as the *foreign language effect* (FLE). Note that the two aspects or main features of the FLE (i.e., the reduced emotionality and the increased rationality) are not unrelated. Rather, common explanations of the FLE propose that the reduced emotionality in an L2 is, indeed, a key mechanism of the increased rationality (but also see the following for additional explanations of the FLE, e.g., Bialek et al., 2020; Geipel et al., 2015; Shin & Kim, 2017). Specifically, reduced emotional reactivity has been suggested to allow for more rational thinking, compared to an L1 context where the relatively larger influence from emotions can lead to more emotional decisions. The *dual-route processing framework* proposes two distinct processes for thinking: System 1, which is characterized by fast, heuristic, and emotional thinking; and System 2, which is slower, more deliberate and rational thinking (e.g., Capraro, 2024; De Neys & Pennycook, 2019; Kahneman, 2003). In the FLE literature, the more automatic System 1 type of thinking has been more closely linked to the more automatic L1, while the less automatic L2 is more often associated with the slower, more deliberate and less emotional System 2 type thinking (e.g., Keysar et al., 2012; Circi et al., 2021; but also see Geipel et al., 2015, who suggested that the FLE in, for example, moral judgments, stems from a reduced access to normative knowledge in a foreign language).

However, precisely how these two forces interact, and how the underlying cognitive mechanisms and emotional processes are affected, is not yet fully understood. Various aspects have been suggested to modulate the FLE, such as immersion in one's L2 country (e.g., Čavar & Tytus, 2018), or the modality in which the stimulus materials are presented (e.g., Brouwer, 2019, 2021). Most relevant for this study is the notion that linguistic similarity may also modulate the FLE. This was first proposed by Dylman and Champoux-Larsson (2020) who

found that L2 languages that were highly influential (such as the “lingua franca” English in Sweden), or which were linguistically very similar to participants’ L1 (such as Swedish for L1 Norwegian speakers or Norwegian for L1 Swedish speakers) did not produce an FLE.

Dylman and Champoux-Larsson thus proposed that typologically similar languages, through shared linguistic structure, grammar, and cognates, could activate the emotional resonance of the L1, thereby resulting in the FLE being attenuated or even cancelled out completely.

Based on this notion, languages that are linguistically more distant to one’s L1 should result in a larger FLE compared to languages which are more similar to the L1. However, there are some inconsistencies in the literature. For example, in their meta-analysis, Stankovic et al. (2022) explored linguistic similarity as a moderating variable, and found that it did not significantly predict an FLE in the selected studies, while Circi et al.’s (2021) meta-analysis did find an effect of linguistic similarity. These ad hoc analyses, however, include experiments that have been carried out without the role of linguistic similarity in mind. As such, there may be variability between the compiled studies, and the analyses will be limited to the language pairs tested. This may be problematic given the wider issue of an overreliance of a selected few languages both in L2 acquisition research generally (e.g., Bylund et al., 2024) and in the FLE literature specifically (e.g., Bylund & Athanasopoulos, 2025). Thus, the current study aimed to empirically investigate the role of linguistic similarity on the magnitude of the FLE in decision making and emotional resonance in an L2 by investigating a diverse number of L2s which allows us to systematically categorise participants based on linguistic similarity between their L1 and L2. To this end, two experiments were conducted examining participants’ responses to three different decision-making tasks: The Asian disease problem, investigating framing effects; a moral dilemma task, investigating moral judgments; and the Cognitive reflection test, measuring the ability to override intuitive but incorrect responses and instead engage in more deliberate thinking processes. All three of these tasks

have been employed in previous studies investigating the FLE; the Asian disease problem and the moral dilemma, in particular, are common tasks in the FLE literature (see, e.g., Circi et al., 2021; Maschio et al., 2022; Stankovic et al., 2022). The Cognitive reflection test has also been used in the FLE literature (e.g., Del Maschio et al., 2025), if to a lesser extent than the other two tasks. Milczarski et al., (2023), for example, found no effect of language on the Cognitive reflection test. However, the populations tested consisted mainly of L1 Polish and L2 English speakers, and occasionally L2 German, French, or Spanish, all of which typologically similar languages (relatively speaking). Thus, for the sake of this exploratively natured study, the Cognitive reflection test was added to the list of tasks. In addition to the three decision-making tasks, we also asked the L2 participants to report their level of emotional resonance in their L2, or more accurately, their reduction of emotional resonance in their L2 compared to their L1. This was measured using the Reduced Emotional Resonance in LX (RER-LX, Toivo et al., 2023).

In Experiment 1, three groups of participants responded to these three decision-making tasks in English, where one group were L1 speakers of English, and the remaining participants had English as an L2. These latter participants were divided into two groups of L2 speakers, based on the linguistic similarity between their L1 and English, resulting in one group of L2 speakers with an L1 which was deemed more similar to English (the L2 Close group), and one group with an L1 which was more distant from English (the L2 Far group). This partly resembled the categorisation of linguistic similarity that was used by Stankovic et al. (2022).

Experiment 2 aimed to replicate (with a few methodological differences) Experiment 1 using another language as the target language, namely Swedish. Similarly to Experiment 1, Experiment 2 asked three groups of participants (L1 Swedish, L2 Close, and L2 Far) to respond to the three decision-making tasks (Asian disease problem, moral dilemma, and the

Cognitive reflection test), as well as fill out the RER-LX scale. In combination, the two experiments provide data on the role of linguistic similarity in the FLE (using three different decision-making tasks) and emotional resonance, in two different target languages. As such this study attempts to more directly test the role of linguistic similarity (as initially put forward by Dylman and Champoux-Larsson, 2020), using L2 speakers from various language backgrounds. Based on previous indications, we tentatively hypothesised that L2 Far groups would demonstrate a greater FLE and show more utilitarian responses than the L2 Close groups, and that both these groups would make more utilitarian and rational choices than groups performing in an L1 context.

Linguistic similarity was categorised based on how linguistically close or distant the participants' L1 was to their L2 English in Experiment 1, and L2 Swedish in Experiment 2. This categorisation was based on multiple facets, including typology, orthography, language family, and indices of linguistic distance based on estimates of how easy a language is to learn (Chiswick & Miller, 2005). We also considered participants' own perception of the linguistic distance between their L1 and L2, so-called psychotypology, as there is evidence that perceived similarity between one's languages may be as, if not more, important when it comes to transfer between languages (e.g., Kellerman, 1986; Sayehli, 2013). A more thorough description of the categorisation is provided in the Methods section of each experiment.

Thus, the overall aim of this study was to investigate whether linguistic distance modulates the reduction of emotionality in the L2, thereby differentially affecting the FLE. Specifically, we attempted to examine the effect of language context (L1 vs. L2) and linguistic similarity, that is, potential differences in the FLE between linguistically similar (close) and distant (far) languages.

Experiment 1: English

Methods

Participants

Three versions of the experiment were set up on Prolific Academic, all advertised and pre-screened for location (U.K.), age (minimum 18), bilingualism (first language plus one other), and English proficiency (fluent). One version was set up to recruit (i.e., pre-screen for) speakers of any first language (including English). This was expected to recruit participants to all three groups (L1 English, L2 English-Close, L2 English-Far) but would likely attract more participants for L1 English. A second was set up to recruit speakers of any first language other than English. This was expected to populate the L2 English-Close and L2 English-Far groups. Finally, a third version was set up during recruitment which limited participants to those with first languages that would be categorised as linguistically far from English. This was because the majority of participants whose L1 was not English spoke an L1 that was linguistically close, such as Italian and French, and group numbers were unbalanced. Data from three L1 English speakers were retained despite stating during the experiment that they had no second language, as bilingualism was less important a criterion for this group. In total, 104 participants completed the study (mean age = 35 years, $SD = 11$; 41 men, 61 women, 2 other/preferred not to say). Final numbers were 31 for the L1 English group, 37 for L2 English-Close, and 36 for the L2 English-Far.

English language proficiency was measured using separate self-report questions (on 10-point Likert scales) for general proficiency, reading, writing, speaking and listening respectively, as well as age of acquisition (AoA), that is, the age at which they started acquiring English. Table 1 reports the English language proficiency levels and AoA for the participants in Experiment 1. Further, the mean age of moving to the UK was similar between the L2 Close ($M = 25.9$ years, $SD = 8.3$) and the L2 Far group ($M = 23.8$ years, $SD = 8.8$). A one-way between-subjects ANOVA on the average language proficiency levels revealed a significant difference between the three groups, $F(2, 88) = 5.46$, $p = .006$, $\eta^2 = .11$. Tukey's

posthoc test revealed that the L1 group was significantly more proficient in English than the L2 Far group, $p = .004$, but no significant difference was found between the L1 and L2 Close ($p = .427$), or the L2 Close and L2 far group ($p = .068$). A one-way between-subjects ANOVA on the Age of Acquisition (AoA) of English found a significant difference between the three groups, $F(2, 88) = 35.4$, $p < .001$, $\eta^2 = .45$, and Tukey's posthoc test revealed, unsurprisingly, that the L1 group acquired English significantly earlier than both the L2 Close and the L2 Far group (both $p < .001$). The L2 Close and the L2 Far group did not differ from each other in terms of the age at which they started acquiring English, $p = .092$.

Table 1

English language proficiency and Age of Acquisition (AoA) of participants in Experiment 1.

	L1 English	L2 English (Close)	L2 English (Far)
	Mean (SD)	Mean (SD)	
General proficiency	9.61 (0.8)	9.1 (1.0)	8.8 (1.0)
Reading	9.65 (0.8)	9.59 (0.7)	9.1 (1.0)
Writing	9.4 (0.9)	9.2 (1.4)	8.7 (1.1)
Listening	9.5 (1.0)	9.4 (0.8)	8.9 (0.9)
Speaking	9.5 (1.1)	9.0 (1.0)	8.4 (1.3)
Mean	9.5 (0.8)	9.3 (0.8)	8.8 (0.9)
English AoA	2.0 (1.9)	9.0 (4.6)	7.1 (2.9)

Note. Means and standard deviations for the English proficiency measures (measured on 10-point Likert scales) and Age of Acquisition in years.

Stimulus materials

The main materials used in Experiment 1 consisted of three different tasks. First, a version of the Asian disease problem (taken from Dylman & Champoux-Larsson, 2020) was presented in English. Participants were told that an Asian disease was expected to kill 600,000 people, and they should choose one of two options (Program A and Program B). Two versions were presented, either a 'gain frame' context or 'loss frame' context. In the gain frame

context, Program A would save 200,000 lives, and Program B offered a one-third probability that all 600,000 would be saved (and a two-thirds probability that no-one would be saved). In the loss frame context, Program A would result in the deaths of 400,000, and Program B offered a one-third probability that no-one would die (and two-thirds that all would die). Second, a version of the footbridge moral dilemma was presented in English, and participants reported their response (i.e., how likely there are to pull the lever), emotionality (how emotional the dilemma made them feel), and comprehension (how well they understood the dilemma) on a continuous 6-point Likert-scale. Third, a Cognitive reflection test was presented, comprising three open-ended questions presented in English. The questions are designed to measure how well people can override initial intuitive (but incorrect) responses and instead engage in more deliberate thinking. Each item typically has an answer that feels immediately obvious but is actually wrong, prompting the participant to slow down and reflect. Correct scores are associated with a greater tendency to think analytically rather than relying on gut impressions. See Appendix A for all three items in the Cognitive reflection test.

Furthermore, the 15-item RER-LX scale was administered in order to measure the hypothesised reduced emotional resonance in bilinguals' later learnt language (Toivo et al., 2024). This was used with a small modification, whereby instances in the original scale of "my LX" were changed to "English" (e.g., the first item "I feel less emotional when using my L2 than when using my L1" was changed to "I feel less emotional when using English than when using my L1"). This was done in order to make explicit that the comparison was between their L1 and their L2 English.

Additionally, two attention check questions were presented, asking participants to select a specific number on a 10-point scale. This was done to ensure that the participants were continuously paying attention, to ensure the validity of the responses to the main tasks,

and participants who failed to comply with the instructions on the control questions were subsequently excluded from analysis.

Finally, background questions, measuring participants language proficiency levels in English, their L1a, AoA, language use, and general demographics such as age and gender, were asked.

Procedure

The experiment was presented via Qualtrics, and distributed and completed online. Following information and informed consent, the participants were presented with the three main tasks (Asian disease problem, moral dilemma, and the Cognitive reflection task). For the Asian disease problem, each participant was presented with either the Gain-frame or the Loss-frame version. The order of the three tasks as well as the different framing versions (for the Asian disease problem) was evenly randomized across all participants. The order of the two options for the dichotomous options (in this case the A vs. B in the Asian disease problem) was also randomized for every participant, as was the order of the three questions in the Cognitive reflection task. All participants were asked to respond to all three questions in the Cognitive reflection task. The two control questions were presented after the moral dilemma and the Cognitive reflection task respectively. Finally, the language and background questions were presented.

Linguistic Distance (categorisation)

The participants self-reported first languages were inspected for linguistic distance, or similarity, in relation to English. First, the participants who had reported English as an L1 were automatically placed in the L1 English group. For the participants who had self-reported English as an L2, an overview of their language profile was jointly inspected by the authors. Each reported (non-English) first language was categorised as being either close in similarity or distance to English (Close) or more dissimilar/distant to English (Far) based on linguistic

attributes such as orthography and language family group. To illustrate, languages such as Spanish and German both belong to the Indo-European language family (as does English) and also share the same orthography (except for a few letters). In contrast, languages such as Japanese or Arabic belong to other language families and their writing systems differ from English. Thus, the former languages (i.e., Spanish and German) were categorised as L2 Close, whereas the latter examples (i.e., Japanese and Arabic) were categorised into the L2 Far group.

In a second step, we compared these classifications against Chiswick and Miller's (2004) index of Linguistic distance, which provides a quantitative measure of the linguistic distance between English and various other languages. A strong positive correlation was observed between the two measures, $r(40) = .90$, $p < .001$, which confirmed the validity of our classification system. See Appendix B for a full list of the reported languages and their categorisations. Note that the number of languages does not correspond to the number of L2 speaking participants as multiple participants indicated the same language as their L1.

This classification resulted in the following final numbers: L1 English = 31, L2 English Close = 37, L2 English Far = 36, total L2 = 73.

Linguistic Distance (self-rating)

In addition to the above classification, we also wanted to explore the participants' own perceptions of the linguistic similarity between their L1 and English. This was measured on a 10-point Likert scale, with higher ratings indicating that the participants perceived their L1 to be more distant/dissimilar to English. A Mann-Whitney U test found that participants in the L2 English (Far) group judged their L1 to be more different from English than participants in the L2 English (Close) group, $M_{Far} = 8.2/10$; $M_{Close} = 7.5/10$, $U(73) = 488.5$, $p = .046$, rank biserial correlation $r = 0.27$. The data were not normally distributed (Shapiro Wilk tests $p < .001$ and $p = .001$ respectively). Thus, the self-rated measures of linguistic similarity

confirmed our categorisation, and suggested that the participants in the Close group assessed their L1 to be more similar to English than did the participants in the Far group. However, it is noteworthy that the difference is small in real terms (less than one point on the ten-point scale), which may reflect either that the distance between Far and Close is small, or that the question is generally a difficult one to introspect upon.

Results

All statistical analyses were performed using JASP version 0.19.3 (JASP Team, 2024).

Asian Disease Problem

All three groups exhibited a significant framing effect (i.e., influence of the Loss vs. Gain framing) in the Asian Disease problem (see Table 2), with higher proportions of participants choosing Program A in the Gain frame than the Loss frame: L1 English: $\chi^2(31) = 5.49, p = .019, \phi = .42$; L2 English (Close): $\chi^2(37) = 11.80, p < .001, \phi = .57$; L2 English (Far): $\chi^2(36) = 11.31, p < .001, \phi = .56$, and in total, $\chi^2(104) = 28.132, p < .001, \phi = .52$. In sum, all three groups were powerfully influenced by the way the scenario was framed, whether this was in their first language or second language.

Table 2.

Proportion of responses to the Asian disease problem in Experiment 1.

	Framing	Program	
		A	B
L1 English	Gain	68.75%	31.25%
	Loss	26.67%	73.33%
L2 English (Close)	Gain	76.47%	23.53%
	Loss	20.00%	80.00%
L2 English (Far)	Gain	73.68%	26.32%
	Loss	17.65%	82.35%

Moral Dilemma

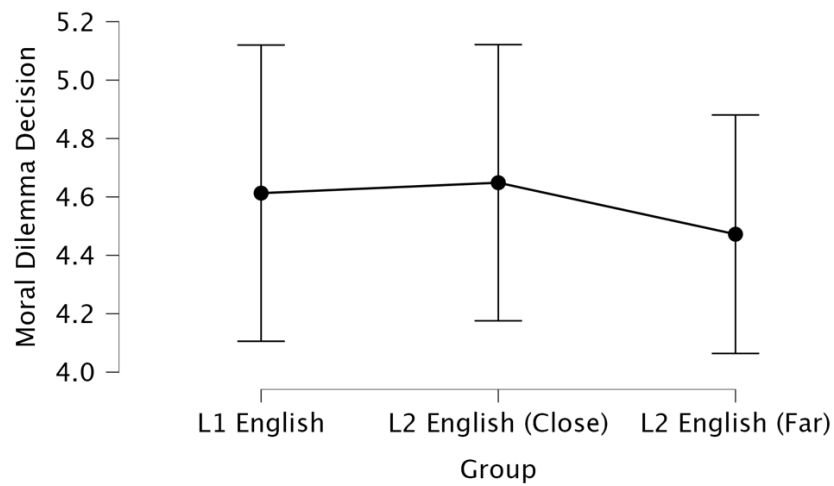
We performed a Kruskal-Wallis test with moral dilemma decision as the dependent variable and group as the independent variable. The test was not significant, $H(2, N = 104) =$

0.95, $p = .62$. Note the means show that participants in all three groups generally preferred the decision to pull the lever (see Figure 1). A Kruskal-Wallis test also found no evidence for a difference between the three language groups on the emotion experienced while contemplating the scenario (see Figure 2), $H(2, N = 104) = 3.25, p = .18$. Finally, a Kruskal-Wallis test also found no evidence for a difference between the three groups on the comprehension of the scenario (see Figure 3), $H(2, N = 104) = 0.48, p = .79$. Participants in all three groups showed near ceiling comprehension level (lowest mean 5.83 out of a maximum of 6). In sum, there was no evidence to support an FLE in decision making as there was no evidence that participants with L2 English would be more likely to pull the lever (i.e., choose the utilitarian option) than participants with L1 English. There was also no evidence to support a greater tendency to choose the utilitarian option when English, the language of the scenario, was a more linguistically distant L2. The absence of any deficit in comprehension in the L2 English groups rules out any misunderstanding as a reason for the absence of an FLE. Finally, an analysis of data only from participants whose L2 was not English found no significant correlations between the likelihood of pulling the lever and either Moral Dilemma Emotionality ($\rho = -.031, p = .80$), or RER-LX mean ($\rho = .02, p = .84^1$), suggesting that neither the strength of emotions experienced while considering one's choice nor the relative emotionality of participants' L1 relative to English was related to one's eventual decision. Additionally, there was no correlation between the two predictors (Moral Dilemma Emotion and RER-LX, $\rho = .15, p = .21$).

Figure 1

Mean ratings of utilitarianism in the footbridge moral dilemma, with 95% confidence intervals.

¹ Additionally, not one correlation analysis between Moral Dilemma Decision and individual items in the RER-LX reached significance (all $ps > .08$).



Note. Higher ratings indicate more likely to pull the lever, i.e., more utilitarian response.

Figure 2

Mean ratings of emotionality of the footbridge moral dilemma, with 95% confidence intervals.

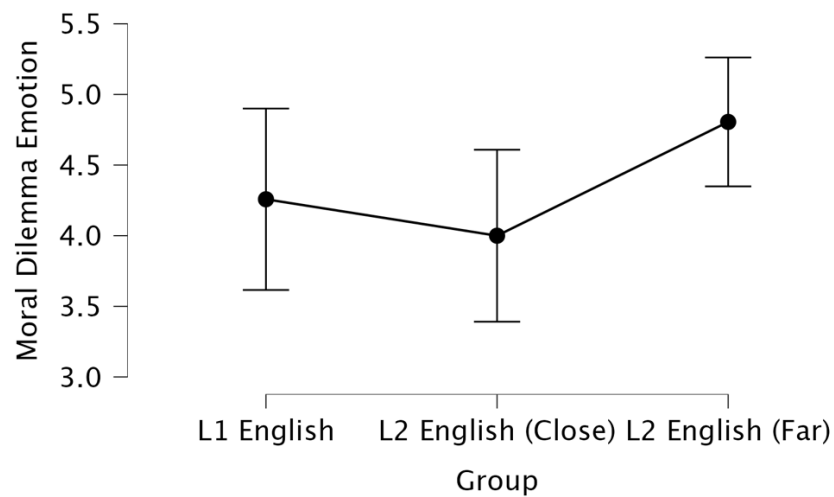
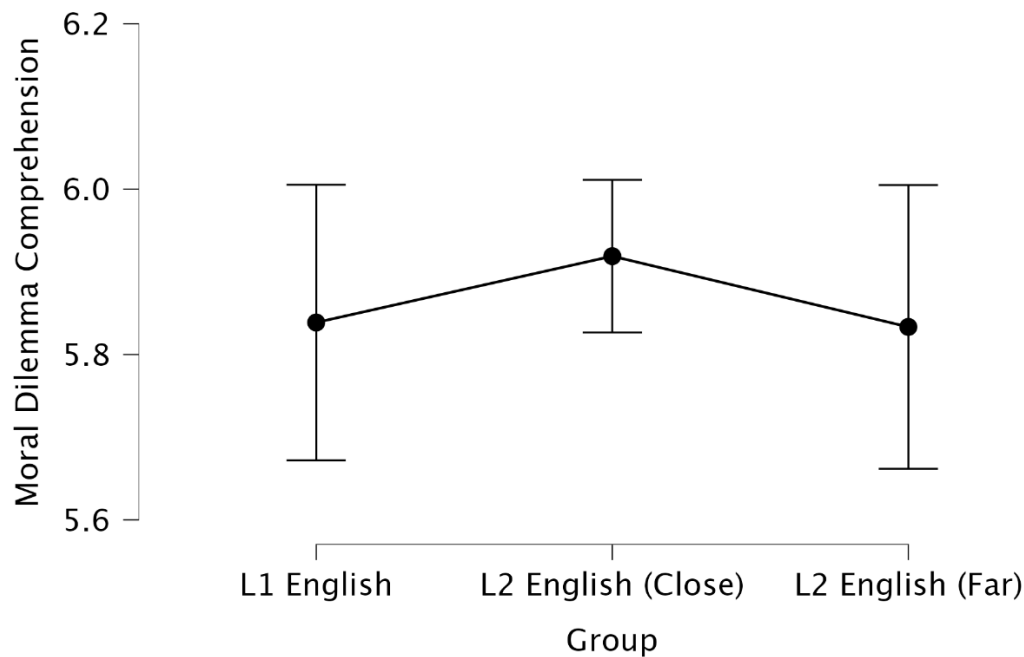


Figure 3

Mean ratings of comprehension of the footbridge moral dilemma, with 95% confidence intervals.



Cognitive reflection tasks

A generalized linear mixed-effects model with a binomial error distribution and logit link function was used to examine the effects of Language group (L1, L2-Close, L2-Far) and task (batball, lilypad, widget) on response accuracy. The model included fixed effects of Language, task, and their interaction, and a random intercept for participants. Likelihood ratio tests indicated that the main effect of Language was not significant, $\chi^2(2) = 1.39, p = .499$, suggesting that overall accuracy did not differ reliably between the three language groups. In contrast, there was a significant main effect of task, $\chi^2(2) = 34.029, p < .001$. The Language $\times$ task interaction was not significant, $\chi^2(4) = 3.247, p = .517$, suggesting that the pattern of task difficulty was similar across language groups.

Estimated marginal means on the response scale showed that accuracy was lowest for the batball item (estimated probability: L1 = .28; L2-Close = .24; L2-Far = .27) and highest for the widget item (estimated probability: L1 = .65; L2-Close = .76; L2-Far = .84). Accuracy on the lilypad item fell between these extremes (estimated probabilities: L1 = .53; L2-Close = .76; L2-Far = .87).

Overall, the model indicates strong differences in task difficulty, with batball being the most challenging and widget the easiest. However, there was no evidence that language background influenced performance overall or moderated the effect of task. In other words, no effect of language (or linguistic similarity) was found for performance on the Cognitive reflection task in this sample.

RER-LX

For the two L2 groups (close and far), we also investigated the relative emotionality of participants' L1 when this was not English to L2 English, as function of linguistic distance, using the RER-LX scale. Overall, participants' whose L1 was linguistically Far from their L2 English scored higher on the RER-LX than participants whose L2 was linguistically Close to English, $M_{Far} = 3.99/6$; $M_{Close} = 3.3/6$, $t(71) = 2.77$, $p = .007$, $d = 0.65$. This suggests that participants judged their L1 to be more emotionally resonant relative to English when that L1 was also linguistically more distant from English.

Discussion

Contrary to the expansive FLE literature, Experiment 1 did not find any effect of language context in any of the three main tasks, namely, the Asian disease problem, the trolley moral dilemma, and the Cognitive reflection task. While this contradicts the well-observed FLE, there may be a couple of possible explanations. The first conceivable explanation concerns the specific population of L2 speakers, where both the L2 Close and the L2 Far groups consisted of highly proficient L2 English speakers (lowest reported self-rated mean English proficiency = 8.8 on a 10-point scale). While meta-analyses show that language proficiency as such does not seem to modulate the FLE (see e.g., Del Maschio et al., 2022), there is another important point to observe in this population, namely their AoA of English. While both L2 groups reported moving to the UK (their L2 country) in their mid-20s, their AoA of English was considerably younger (L2 Close $M = 9$ years; L2 Far $M = 7$ years). Thus,

it could be that English was acquired in a manner more consistent with first languages than later-learned second-languages. However, it is difficult to know where to draw the line between these two possibilities. We interpret these data as suggesting that the conditions under which an FLE can be demonstrated are perhaps more restricted than hitherto supposed.

A few additional methodological aspects also need to be raised. The moral dilemma tasks that have usually been used in the moral foreign language effect (mFLE) literature have typically used so-called personal dilemmas instead of impersonal dilemmas. In personal dilemmas, the reader is faced with a scenario whereby they need to actively carry out an action to kill one person (such as pushing a person in front of a train) whereas the action in impersonal dilemmas is more passive (such as pulling a lever to redirect a train to another track). Even if the outcome of the action in both versions is the same, studies have found that the active action is more difficult and emotional compared to the passive action (e.g., Christensen et al., 2014), which is why most studies on the moral FLE use personal rather than impersonal moral dilemmas (e.g., Del Maschio et al., 2022) even if there exist studies that have successfully used both (see Geipel et al., 2015, who specifically compared the trolley and footbridge moral dilemmas in an FLE context). In this experiment, we opted for the (impersonal) trolley problem over the (personal) footbridge problem. This decision was based on a pilot study where qualitative feedback from several participants cast doubt on the internal consistency of the footbridge problem. Reflections in relation to the internal consistency included the relevance of the man being heavy (insinuating a fat-shaming discourse, i.e., “Why is it explicitly stated that the man to be pushed is heavy?”, “Am I expected to push him because he is heavy?”), which tied into other comments related to the likelihood of stopping a train by pushing a person in front of it (i.e., “If the man were less heavy, would pushing him in front of the train still stop the train?”, “How likely is it that a train could be stopped by a person pushed in front of it, irrespective of said persons

weight/size?"). Notwithstanding, the internal consistency of the specific moral dilemmas used in FLE studies remains an empirical question for now.

Relatedly, most studies on the moral FLE provide a dichotomous choice (with a few exceptions, including Čavar & Tytus, 2018; and Geipel et al., 2015) whereby the participant can either choose the deontological or the utilitarian option (e.g., "Will you pull the lever/push the person?" YES/NO). In contrast, Experiment 1 used a continuous 6-point Likert scale where participants indicated how likely they were to pull the lever, ranging from "definitely no" to "definitely yes", which may have diluted the responses.

While these two methodological points (using the impersonal trolley problem and asking participants to respond on a dichotomous scale) are unlikely to have drastically altered the results, it is possible that the combination of both these points may have contributed to the lack of FLE found in the moral dilemma. However, even if this could explain the lack of FLE for the moral dilemma task, it cannot account for the lack of FLE in the other two tasks used, and so it is more probable that other factors (such as L2 proficiency, AoA, or immersion) can explain the lack of a FLE in this sample.

Interestingly, despite finding no effect of language in any of the three decision making tasks, there was a clear effect of linguistic similarity on reduced emotional resonance in L2, as measured by the RER-LX scale (Toivo et al., 2023), where the participants with an L1 that was more linguistically distant to their L2 reported a larger reduction of emotional resonance in their L2 compared to the L2 speakers with a linguistically more similar L1. This suggests that even when the L2 (for various potential reasons, such as high proficiency, early AoA, or immersion) does not result in a more rational decision-making style, there may still be an effect of linguistic similarity on the emotional resonance experienced in the L2.

Experiment 2: Swedish

Experiment 2 aimed to replicate (with a few modifications) Experiment 1 in a second target language, namely Swedish. In addition, a few methodological modifications were made to rule out the impact of specific methodological aspects affecting the results in Experiment 1. Specifically, Experiment 2 used the personal Footbridge moral dilemma, instead of the impersonal Trolley problem, and provided a dichotomous response option (Yes vs. No) instead of the continuous 6-point Likert scale used in Experiment 1. Apart from these points, all other aspects remained the same.

Methods

Participants

Participants were recruited online via expats in Sweden groups on social media. A brief description of the study together with a link to the study (where more thorough information was provided) was posted in several such groups. For the L1 Swedish group, participants were recruited during a lecture at a Swedish university, where a link to the study was presented. The study was subsequently completed online.

A total of 145 participants completed the study (mean age = 43 years, SD = 11.4; 25 men, 119 women, 1 other/preferred not to say). See Table 3 for participants' language proficiency levels and age of acquisition of Swedish. A one-way between-subjects ANOVA on the average language proficiency levels revealed a significant difference between the three groups, $F(2, 139) = 42.1$, $p < .001$, $\eta^2 = .38$. As would be expected, Tukey's posthoc test revealed that the L1 group were significantly more proficient in Swedish than both the L2 Close and L2 Far group, both $p < .001$. No significant difference was found between the L2 Close and L2 far group, $p = .868$. As for AoA of Swedish, a one-way between-subjects ANOVA found a significant difference between the three groups, $F(2, 139) = 316$, $p < .001$, $\eta^2 = .82$, and Tukey's posthoc test revealed that the L1 group acquired Swedish significantly

earlier than both the L2 Close and the L2 Far group (both $p < .001$). The L2 Close and the L2 Far group did not differ from each other on AoA of Swedish, $p = .738$.

Table 3

Swedish language proficiency and Age of Acquisition (AoA) of participants in Experiment 2.

	L1 Swedish	L2 Swedish (Close)	L2 Swedish (Far)
	Mean (SD)	Mean (SD)	
General proficiency	9.8 (0.7)	7.4 (1.8)	7.1 (1.7)
Reading	9.1 (1.3)	8.3 (1.4)	7.8 (1.4)
Writing	8.9 (1.2)	6.9 (1.8)	6.9 (1.8)
Listening	9.1 (1.2)	7.4 (1.9)	7.5 (1.5)
Speaking	9.1 (1.0)	7.0 (1.9)	7.0 (1.8)
Mean	9.2 (0.9)	7.4 (1.6)	7.2 (1.4)
Swedish AoA	0.4 (1.5)	29.4 (11)	30.7 (3)

Note. Means and standard deviations for the Swedish proficiency measures (measured on 10-point Likert scales) and Age of Acquisition in years.

All Swedish ethical codes, laws, regulations, and guidelines for research involving humans were followed. Written consent was collected from all participants prior to study participation. A detailed explanation of the study (e.g., descriptions of the data to be collected and the study procedure) was given to all participants, and they were explicitly informed that all data would be anonymous and that they could withdraw from the study at any time without any consequences or disadvantages.

Stimulus materials

The same materials as used in Experiment 1 were also used in Experiment 2, with a few exceptions. Instead of the trolley moral dilemma, Experiment 2 used the Footbridge moral dilemma (see Appendix A). In contrast to the continuous 6-point scale to measure the response to the moral dilemma in Experiment 1, Experiment 2 measured the response to the moral dilemma in a dichotomous manner, where participants indicated their willingness to

push the man (sacrifice one life to save the lives of five) by choosing either “yes” or “no”.

This dichotomous response is commonly used in the mFLE literature (see e.g., a meta-analysis by Stankovic et al., 2022). All materials used were presented in Swedish. Where applicable, the Swedish versions from Dylman and Champoux-Larsson (2020) were used (this was true for the Asian disease problem and the footbridge moral dilemma). For the three questions on the Cognitive Reflection Task, the questions were translated and back-translated by two independent assessors (one of whom a balanced L1 English-L1 Swedish bilingual).

Procedure

The same procedure as Experiment 1 was used.

Linguistic Distance (categorisation)

For Experiment 2, the same classification system was used as in Experiment 1, but this time taking into consideration the linguistic distance/similarity between Swedish and each non-Swedish language reported by the participants. To the best of our knowledge, no equivalent index for linguistic distance between Swedish and other languages exists. Given the high correlation between Chiswick and Miller’s index for English and our classification system in Experiment 1, this process of categorisation was assessed as reasonable and valid for Experiment 2. This classification resulted in the following final numbers: L1 Swedish = 77, L2 Swedish Close = 44, L2 Swedish Far = 21, total L2 = 65.

Linguistic Distance (self-rating)

As in Experiment 1, in addition to the linguistic distance categorization, we also explored the participants’ own perceptions of the linguistic similarity between their L1 and Swedish. This was again measured on a 10-point Likert scale, with higher ratings indicating that the participants’ perceived their L1 to be more distant/dissimilar to Swedish. A Mann-Whitney U test found that participants in the L2 Swedish (Far) group judged their L1 to be more different from Swedish than participants in the L2 Swedish (Close) group, $M_{Far} =$

8.6/10, $SD = 2.3$; $M_{Close} = 6.1/10$, $SD = 2.9$, $U(65) = 232.5$, $p < .001$, rank biserial correlation $r = 0.15$. The data were not normally distributed (Shapiro Wilk test $p = .014$).

Results

All statistical analyses were performed using JASP version 0.19.3 (JASP Team, 2024).

Asian Disease Problem

For the L1 Swedish group, a significant effect of framing was found: When presented with the gain-frame version, 78% chose Program A, but when presented with the loss-frame version, only 22% chose Program A, $\chi^2(1, N = 77) = 23.93$, $p < .001$, $\phi = 0.56$. A significant effect of framing was also found for the L2 Close group: When presented with the gain-frame version, 85 % chose Program A, but when presented with the loss-frame version, only 37.5 % chose option A, $\chi^2(1, N = 44) = 10.18$, $p = .001$, $\phi = 0.48$. In contrast, no significant effect of framing was found for the L2 Far group: When presented with the gain-frame version, 36.4 % chose option A, and when presented with the loss-frame version, 20 % chose option A, $\chi^2(1, N = 21) = 0.69$, $p = .407$. In sum, an FLE was found, but only for the L2 Far group. As such, there was an effect of linguistic similarity on the FLE in the framing effect as measured by the Asian disease problem, see Table 4 for descriptives.

Table 4.

Proportion of responses to the Asian disease problem in Experiment 2.

	Framing	Program	
		A	B
L1 Swedish	Gain	78.05%	21.95%
	Loss	22.22%	77.78%
L2 Swedish (Close)	Gain	85.00%	15.00%
	Loss	37.50%	62.50%
L2 Swedish (Far)	Gain	36.36%	63.64%
	Loss	20.00%	80.00%

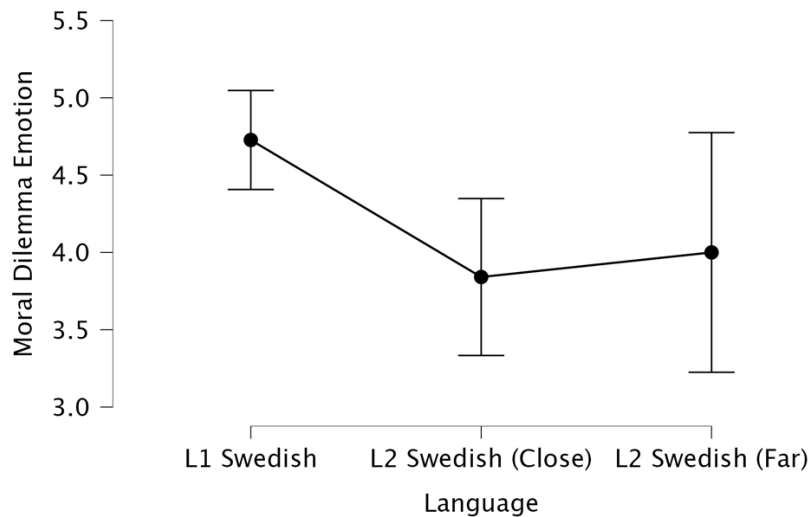
Moral Dilemma

A chi-square test for contingency tables was conducted to investigate the association between the L1 group and the combined L2 group (Close and Far combined) and their responses on the moral (footbridge) dilemma (Yes/Utilitarian vs. No/Deontological). A significant effect of language on the choices the participants made was found: a larger percentage of participants chose the utilitarian option in the L2 groups (utilitarian = 29.4 %) compared to the L1 group (utilitarian = 14.3 %), $\chi^2(1, N=145) = 4.92, p = .027, \phi = 0.184$. When only looking at the L2 groups (Close vs. Far), the chi-square test for contingency table was no longer significant, $\chi^2(1, N = 65) = 1.18, p = .278$. In other words, while a general FLE was found, there was no effect of linguistic similarity on the magnitude of the FLE.

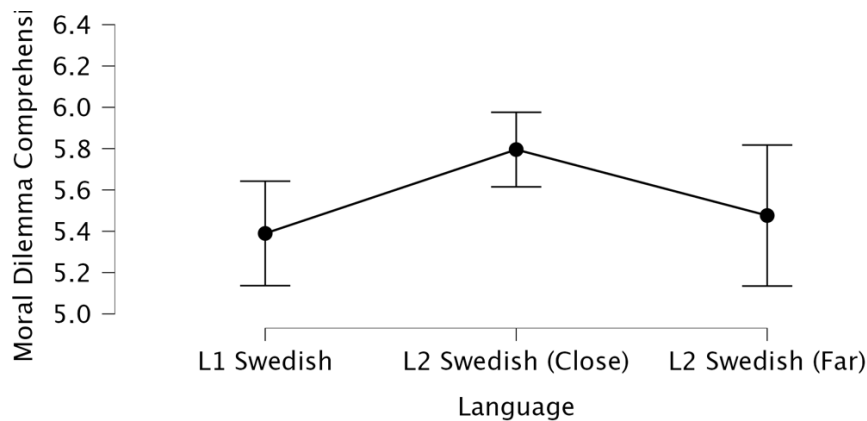
A Kruskal Wallis test found a significant effect of language on the rating of emotionality experienced from reading the moral dilemma (see Figure 4), $H(2, N = 142) = 9.42, p = .009$. Dunn's post-hoc tests applying the Bonferroni correction revealed that the L1 group ($M = .91, SD = .99$) reported significantly higher emotionality compared to the L2 Close group, $p = .012, r_{rb} = .31$, but the difference between the L1 group and the L2 Far group was not significant, $p = .20, r_{rb} = .25$. Finally, there was no difference between the L2 Close and L2 Far group, $p = .72, r_{rb} = .05$. Finally, a one-way between-subjects ANOVA revealed no significant effect of language on the comprehension of the moral dilemma itself, $H(2, N = 142) = 5.66, p = .059$.

Figure 4

Mean ratings of emotionality of the footbridge moral dilemma.

**Figure 5**

Mean ratings of comprehension of the footbridge moral dilemma.



Cognitive reflection tasks

A generalized linear mixed-effects model with a binomial error distribution and logit link function was used to examine the effects of Language group (L1, L2-Close, L2-Far) and task (batball, lilypad, widget) on response accuracy. The model included fixed effects of Language, task, and their interaction, and a random intercept for participants. Likelihood-ratio tests on the fixed effects showed a significant main effect of Language, $\chi^2(2) = 21.94$, $p < .001$, and a significant main effect of Item, $\chi^2(2) = 14.37$, $p < .001$. The Language $\times$ Item interaction was not significant, $\chi^2(4) = 1.75$, $p = .782$, indicating that Language differences were similar across Items. We conducted three prespecified pairwise contrasts for

the three levels of Language (Holm adjusted). Both L2 groups outperformed L1: L2-Close > L1 by 41.9 percentage points (pp), 95% CI [29.4, 52.6], median odds ratio (OR) = 6.31, $p_{adj} < .001$; L2-Far > L1 by 30.6 pp, 95% CI [13.4, 45.9], OR = 3.95, $p_{adj} < .001$. The difference between the L2-Close and the L2-Far group was non-significant: 11.3 pp, 95% CI [-8.9, 31.6], OR = 1.62, $p_{adj} = .277$.

Across the three different tasks, both the L2-Close and the L2-Far group showed substantially higher probabilities of a correct response than the L1. In other words, a significant FLE was found for both L2 groups in comparison to the L1 Swedish group, but there was no effect of linguistic distance on the magnitude of the FLE.

RER-LX

We investigated the relative emotionality of participants' L1 when this was not Swedish to L2 Swedish, as a function of linguistic distance, using the RER-LX scale. No significant difference was found between the L2 Close and L2 far group on their mean RER-LX responses, $M_{Far} = 3.92/6$, $SD = .94$; $M_{Close} = 4.01/6$, $SD = .96$, $t(56) = 0.59$, $p = .556$. This suggests that linguistic similarity did not affect participants' ratings of whether their L1 was more emotionally resonant relative to their L2 Swedish.

Discussion

Generally speaking, Experiment 2 found a classic FLE across all three decision-making tasks, establishing that an effect of language can be observed in the tested population. For the Cognitive reflection test, the participants in both L2 groups (i.e., both the Close and the Far group) outperformed the L1 group, but no difference was found between the L2 Close and Far group. For the Asian disease problem, a framing effect was found for the L1 group, and the L2 Close group, but not for the L2 Far group. The classic FLE was thus found, but only for the group with an L1 that was more different from their L2 Swedish. As such, we found an effect of linguistic similarity on the FLE in the Asian disease problem. Finally, an

effect of language context was found for the decisions made when faced with the footbridge moral dilemma in that both L2 groups were more likely to choose the utilitarian option compared to the L1 group. No differences were found between the L2 Close and L2 Far group. Similarly, the L1 group reported higher emotional reactivity towards the moral dilemma itself compared to both the L2 Close and L2 Far groups, but the two L2 groups did not differ from each other on emotionality. In other words, we found a classic FLE for the moral dilemma, both in terms of emotionality and the choices made, but we did not observe an effect of linguistic similarity on the magnitude of the FLE.

Contrary to Experiment 1, Experiment 2 found no effect of linguistic similarity on the RER-LX ratings. Qualitative feedback from these participants, however, revealed that multiple participants had a hard time responding to several of the items on the RER-LX scale due to not having found themselves experiencing the particular situation described in the statements (specifically mentioning several items, including “I find it easier to talk about sex in my L1 than in Swedish”). This could have potential implications for the use of the RER-LX, indicating that a degree of experience with emotional situations in one’s L2 is necessary to be able to adequately consider the statements. In the current experiment, we used a forced response approach to maximise data collection, but future studies may need to make all responses optional, to allow participants to leave individual items unanswered. The RER-LX scale may then need to be re-evaluated, in order to further investigate the degree to which a subscale of the items is sufficient to generate analysable and informative data.

General discussion

Across two experiments, we investigated the role of linguistic similarity on the magnitude of the foreign language effect in decision making and emotional resonance. In Experiment 1, using English as the target language, we found no FLE on the Cognitive reflection task, the framing effect in the Asian disease problem, or the moral dilemma. We

did, however, find an effect of linguistic similarity on the reduced emotional resonance in L2, whereby the reduction of emotional resonance was larger for participants whose L1 was linguistically more distant compared to more similar, indicating that larger similarity between the languages may attenuate the reduced emotional resonance.

In Experiment 2, using Swedish as the target language, we generally found an effect of language for all three decision-making tasks, where participants in the L2 context scored higher on the Cognitive reflection task, were less affected by framing in the Asian disease problem, and were more likely to make a utilitarian choice when faced with a moral dilemma, compared to the participants in the L1 context. Generally speaking, we found no effect of linguistic similarity on the magnitude of the FLE (with one exception, where the L2 Close group were affected by the framing in the Asian disease problem whereas the L2 Far group were not). Furthermore, unlike in Experiment 1, we found no effect of linguistic similarity on the RER-LX scale. However, as mentioned in the Discussion section of Experiment 2, this latter result may be explained by the participants in Experiment 2 not being able to relate to several of the situations described in the RER-LX scale, as evident by the qualitative feedback received from the participants. As such, the RER-LX scale may be better suited for participants who have more extensive experience with emotionally varied situations in their L2.

Generally speaking, then, there is some indication to support the notion that linguistic similarity or distance may play a role in the level of emotional reactivity, or rather magnitude of the reduced emotionality experienced in an L2, even if this is not as clearly evident in decision-making tasks.

As for the discrepancy between the two experiments in terms of the prevalence of the effect of language, this may be attributed to differences between the samples. The sample of L2 Swedish speakers in Experiment 2 reported a later AoA compared to the UK sample of L2

English speakers. Interpreting this within the dual-route processing framework, this could mean that using L2 Swedish for the L2 speakers in Experiment 2 is likely more cognitively demanding, which encourages more deliberate System 2 style thinking to a larger extent. This is not observed in the earlier AoA of L2 English speakers in the UK sample, which diminishes the FLE (such as in the Cognitive reflection task, framing in the Asian disease problem, and the degree of utilitarian responses in the moral dilemma). However, when there is a personal emotional element involved (such as the statements in the RER-LX scale), we see an effect of linguistic similarity. As mentioned, this was not observed in the Swedish sample because some of the items described situations that they had not experienced (possibly due to less exposure to their L2 relative to the UK sample's L2 English), as corroborated by qualitative and anecdotal comments from the sample in Experiment 2.

Thus, the L2 English speakers in Experiment 1 acquired English much earlier than the L2 Swedish sample acquired Swedish, and could therefore answer the RER-LX better, because it is based on questions about experience of emotional situations. Relatedly, we found an FLE in the Swedish sample but not UK sample, which could indicate that the L2 Swedish participants find their L2 Swedish more cognitively demanding than the L2 English speakers found their L2 English. This in turn could have led to the L2 Swedish participants being more likely to engage in more deliberate and rational system 2 thinking compared to the L2 English speakers in Experiment 1. However, it could also be that other factors related to differences in AoA and/or L2 proficiency have resulted in differences in L2 emotionality between the two groups. For example, some studies have suggested that the FLE is not due to “increased deliberation but by blunting emotional reactions” (Hayakawa et al., 2017, p. 1387). Caldwell-Harris and Aycicegi-Dinn (2021) also argued that the FLE is due to attenuated emotional responses, but also point out that this is likely due to cognitive load. Either way, the differences between the L2 speakers in Experiment 1 vs. Experiment 2 become clear when

comparing the participants' background. While the L2 participants in both experiments moved to their L2 country during adulthood (on average, the L2 participants in Experiment 1 moved to the UK around the age of 25, while the L2 participants in Experiment two moved to Sweden around the age of 30), there is a significantly larger difference between their respective AoAs. The participants in Experiment 2 started acquiring L2 Swedish around the same time as they moved to Sweden ($M = 30$), whereas the L2 English speakers started acquiring English during childhood, long before moving to the UK ($M = 8$ years), and have likely encountered English more in general (in terms of frequency of exposure) and in more varied contexts, including emotional settings (e.g., through movies, music, travel abroad etc.). Even though AoA is considered particularly relevant for the perceived emotionality of a language (Circi et al., 2021), the literature on AoA effects in the FLE is scarce, and variables such as language proficiency, exposure or use have sometimes been used as a proxy for AoA. If AoA has been specifically studied, it has more commonly been used dichotomously comparing monolinguals with bilinguals, and more nuanced comparisons within bilingual (or L2) populations have been lacking, leading recent meta-analytic studies to disregard AoA as a variable to be specifically investigated (e.g., Stankovic et al., 2022). Studies on emotionality in, for example, heritage speakers, however, have found equal levels of emotional reactivity in heritage speakers' both languages, indicating that the interplay among AoA, proficiency, context of use and emotional reactivity is complex (Caldwell-Harris, 2014; Harris et al., 2006).

This is likely an inevitable effect of using English as the target language. Being a *de facto* lingua franca, it is not uncommon for people all around the world to learn English in school regardless of their prospects of visiting, let alone moving to, an English-speaking country. In contrast, most people (with the exception of special interest) do not learn Swedish as a second language unless they specifically need to, often for migration purposes. As such,

English as a world language presents a special case, which may have implications for its applicability as one of the testing languages in studies investigating the FLE. We elaborate on this aspect in Dylman and Kikutani (in press), and question whether the overreliance on English as a target language in the FLE literature truly reflects general second language processing or whether it is restricted to English, and call for a more inclusive (and in some ways less imperialistic) approach to studying the FLE. A similar point has also been made by Bylund et al. (2024) who highlighted the marked discrepancy in second language acquisition and multilingualism research between the geographical regions where data are typically collected (mainly in monolingual societies in the Global North), and the geographical regions with the highest levels of linguistic diversity and societal multilingualism (such as in the Global South). This argument has also been made more specifically for the FLE literature (Bylund & Athanasopoulos, 2025).

Note that we did not restrict the L2 to one specific language in this study. Instead, we recruited participants with any L1, resulting in a wide range of L1s. This approach has both strengths and limitations. One limitation is that the increased variability may have diluted the effects and made it more difficult to observe clear effects of language context or linguistic similarity. It also has a practical limitation, in that we were unable to investigate these effects in L1 and L2 in the same population (as we could not control for all participants' L1s prior to their participation in the study). This may have led to undesired influence from external variable, such as systematic differences between the L1 and the L2 groups in terms of cultural, or other aspects. Differences in culture have been shown to explain effects that have previously been attributed to differences in language status (e.g. bilinguals vs. monolinguals: Samuel et al., 2018), though differences in language status may also explain effects originally attributed to culture (e.g. Wu & Keysar, 2007). Conversely, however, the methodological strength of including a wide range of L2s makes the results more ecologically valid and

generalisable, as they are not tied to one or two specific instances, but rather incorporate a larger definition of L2 speakers thereby reflecting a more naturalistic setting. This also relates back to Bylund et al.'s (2024) argument mentioned above that “multilingualism needs to be studied in its full diversity” (p. 308).

Another methodological factor to highlight is the classification of similar (Close) and more distant (Far) languages. This dichotomous classification system was deemed suitable for this exploratory study, but future studies may wish to develop a more fine-grained system for measuring linguistic similarity on a continuous scale (e.g., based on the Levenshtein distance). Such a system could then be used to more closely investigate a potential correlation between an index of linguistic similarity and the magnitude of the FLE.

To conclude, this is the first study to directly investigate the role of linguistic similarity on the magnitude of the FLE in decision-making and emotional resonance. These initial findings do not support the notion of linguistic similarity consistently having a significant effect on the magnitude of the FLE in decision-making (in line with Stankovic's findings), but may have an effect on the magnitude of the reduction of emotional resonance in an L2 (as well as in some specific instances in framing effects, the latter consistent with Circi et al.'s findings). This also begs the question of the respective contribution of language similarity vs. cultural aspects that led to an attenuated FLE in Dylman and Champoux-Larsson's Swedish-English, and Swedish/Norwegian-Norwegian/Swedish participants. The seemingly disparate results may seem inconclusive, but they do suggest that the FLE is complex and may depend on various factors, such as, for example, age of acquisition, context of learning, and the linguistic similarity between L1 and L2. The FLE seems to be context dependent and vary between specific task and/or populations, and consistently using English as one of the target languages may obscure the true nature of the FLE, which is clearly not a one-size-fits-all effect. This is in line with the recent null-effects reported by Del

Maschio et al., (2025) who called for a “need to reconceptualize the FLE and its implications on decision-making” (p. 1). While further empirical studies are needed to hone and expand on this to find the exact conditions under which an FLE is observed, including more specifically inspecting the role of emotion processing, we provide another step towards understanding the role of linguistic similarity on the foreign language effect.

Acknowledgments: We wish to thank Faiza Osman for her assistance with and discussions regarding the categorisation of linguistic distance.

Declaration of interest: The authors have no conflicts of interest to declare.

References

- Białek, M., Muda, R., Stewart, K., Niszczoła, P., & Pieńkosz, D. (2020). Thinking in a foreign language distorts allocation of cognitive effort: Evidence from reasoning. *Cognition*, 205. <https://doi.org/10.1016/j.cognition.2020.104420>
- Brouwer, S. (2019). The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *Journal of Multilingual and Multicultural Development*, 40, 865-878. <https://doi.org/10.1080/01434632.2019.1585863>
- Brouwer, S. (2021). The interplay between emotion and modality in the foreign-language effect on moral decision making. *Bilingualism: Language and Cognition*, 24, 223-230. <https://doi.org/10.1017/S136672892000022X>
- Bylund, E., & Athanasopoulos, P. (2025). Familiarity breeds contempt: Decision-making in multilingual South Africa. Conference presentation at ISB15, San Sebastián, Spain.
- Bylund, E., Khafif, Z., & Berghoff, R. (2024). Linguistic and geographic diversity in research on second language acquisition and multilingualism: An analysis of selected journals. *Applied Linguistics*, 45, 308-329. <https://doi.org/10.1093/applin/amad022>

- Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: theoretical implications. *Frontiers in Psychology*, 5, 1055-1055.
<https://doi.org/10.3389/fpsyg.2014.01055>
- Capraro, V. (2024). The dual-process approach to human sociality: Meta-analytic evidence for a theory of internalized heuristics for self-preservation. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspa0000375>
- Čavar, F., & Tytus, A. E. (2018). Moral judgement and foreign language effect: when the foreign language becomes the second language. *Journal of Multilingual and Multicultural Development*, 39, 17-28.
<https://doi.org/10.1080/01434632.2017.1304397>
- Chiswick, B. R., & Miller, P. W. (2005). Linguistic Distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1-11. <https://doi.org/10.1080/14790710508668395>
- 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, 607-607. <https://doi.org/10.3389/fpsyg.2014.00607>
- Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, 28, 1131-1141.
<https://doi.org/10.3758/s13423-020-01871-z>
- Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 43, 1109-1128.
<https://doi.org/10.1037/xlm0000356>

- 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.
<https://doi.org/10.1016/j.cognition.2013.11.010>
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014b). Your morals depend on language. *PloS One*, 9, e94842-e94842.
<https://doi.org/10.1371/journal.pone.0094842>
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28, 503-509.
<https://doi.org/10.1177/0963721419855658>
- Del Maschio, N., Bellini, C., Abutalebi, J. and Sulpizio, S. (2025). Boundaries of the foreign language effect under conditions of uncertainty. *Bilingualism: Language and Cognition*, 1–11. <https://doi.org/10.1017/S1366728925100400>
- 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, 617-630. <https://doi.org/10.1017/S1366728921001012>
- Dewaele, J. M. (2010). *Emotions in multiple languages*. Palgrave Macmillan.
- Dylman, A.S., & Champoux-Larsson, M-F. (2020). It’s (not) all Greek to me: Boundaries of the foreign language effect. *Cognition*, 196.
<https://doi.org/10.1016/j.cognition.2019.104148>
- Dylman, A.S., & Kikutani, M. (forthcoming). Language and emotion. In B. K. Ashdown (Ed.), *A Research Agenda for Cross-Cultural Psychology*. Edward Elgar Press.
- Geipel, J., Hadjichristidis, C., Surian, L. (2015). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS ONE*, 10, e0131529.
<https://doi.org/10.1371/journal.pone.0131529>

- Hayakawa, S., Lau, B. K. Y., Holtzmann, S., Costa, A., & Keysar, B. (2019). On the reliability of the foreign language effect on risk-taking. *Quarterly Journal of Experimental Psychology*, 72, 29-40. <https://doi.org/10.1177/1747021817742242>
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the Foreign-Language Effect on moral judgment. *Psychological Science*, 28, 1387–1397.
<https://doi.org/10.1177/0956797617720944>
- JASP Team (2024). JASP (Version 0.18.3) [Computer software].
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist*, 58, 697-720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Kellerman, E. (1986). An eye for an eye: crosslinguistics constraints on the development of the L2 lexicon. In E. Kellerman & M. Sharwood Smith (Eds.), *Cross-linguistic influence in second language acquisition* (pp. 35-48). Pergamon.
- 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.
<https://doi.org/10.1177/0956797611432178>
- Pavlenko, A. (2006). *Emotions and Multilingualism*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511584305>
- Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study. *Cognitive science*, 42(7), 2313-2341.
- Sayehli, S. (2013). Developmental Perspectives on Transfer in Third Language Acquisition. [Doctoral Thesis, Lund University].

- Shin, H. I., & Kim, J. (2017). Foreign language effect and psychological distance. *Journal of Psycholinguistic Research*, 46(6), 1339-1352. <https://doi.org/10.1007/s10936-017-9498-7>
- Stankovic, M., Biedermann, B., & Hamamura, T. (2022). Not all bilinguals are the same: A meta-analysis of the moral foreign language effect. *Brain and Language*, 227, 105082-105082. <https://doi.org/10.1016/j.bandl.2022.105082>
- Toivo, W., Scheepers, C., & Dewaele, J.-M. (2024). RER-LX: A new scale to measure reduced emotional resonance in bilinguals' later learnt language. *Bilingualism: Language and Cognition*, 27, 434-446. <https://doi.org/10.1017/S1366728923000561>
- Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological science*, 18(7), 600-606.

Appendix A

Materials used in Experiments 1 (and in Experiment 2, but translated into Swedish).

Asian disease problem

Gain-frame version

Imagine that the state is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600,000 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200,000 people will be saved.

If Program B is adopted, there is a one-third probability that 600,000 people will be saved and a two-thirds probability that no people will be saved.

Which of the two programs would you favor?

Loss-frame version

Imagine that the state is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600,000 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 400,000 people will die.

If Program B is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600,000 people will die.

Which of the two programs would you favor?

Trolley moral dilemma (only Experiment 1)

A train is about to run over five people. You have the power to pull a lever and redirected the train to a different track where it would run over one person. There is no time to get anyone off the tracks.

On a scale between 1-6 (1 = Definitely no, 6 = Definitely yes), would you pull the lever?

Footbridge moral dilemma (only Experiment 2)

The footbridge moral dilemma, used in Experiments 2-3 A train is going very fast toward five people stuck on the track. The train has a problem and cannot be stopped, unless a heavy weight is dropped on the track. There is a very heavy man next to you—your only way to stop the train is to push the man onto the track, killing him to save these five people.

Would you push him?

YES NO

Cognitive Reflection Test

1. A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost?

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half the lake?

Appendix B

List of the non-English languages reported in Experiment 1 and their classification.

Language	Categorisation	Number of participants
Afrikaans	Close	1
Croatian	Close	1
Czech	Close	1
Dutch	Close	2
Estonian	Close	1
Finnish	Close	3
French	Close	5
German	Close	3
Hungarian	Close	2
Italian	Close	10
Latvian	Close	1
Lithuanian	Close	2
Polish	Close	2
Portuguese	Close	1
Romanian	Close	2
Arabic	Far	6
Bengali	Far	2
Bulgarian	Far	4
Cantonese	Far	1
Chinese	Far	5
Farsi	Far	1
Greek	Far	3
Hindi	Far	1
Japanese	Far	1
Mandarin	Far	1
Punjabi	Far	1
Russian	Far	1
Tagalog	Far	1
Urdu	Far	2
Tamil	Far	1

Turkish	Far	3
Vietnamese	Far	1
Yoruba	Far	1

Appendix C

List of the non-Swedish languages reported in Experiment 2 and their classification.

Language	Categorisation	Number of participants
Bosnian	Close	1
Croatian	Close	3
Danish	Close	2
Dutch	Close	1
English	Close	23
German	Close	3
Hungarian	Close	3
Italian	Close	1
Polish	Close	2
Portuguese	Close	3
Rumanian	Close	1
Spanish	Close	3
Arabic	Far	4
Dari	Far	1
Greek	Far	2
Hindi	Far	1
Indonesian	Far	1
Persian	Far	2
Russian	Far	1
Serbian	Far	1
Thai	Far	3
Turkish	Far	2
Ukrainian	Far	1