Moral Judgements in a Foreign Language: Expressing Emotions and Justifying Decisions

[a]Andreas Kyriakou

[a,b] Alice Foucart

[a,b,c] Irini Mavrou*

[a] Departamento de Lenguas Aplicadas, Universidad Antonio de Nebrija, Madrid, Spain

[b] Centro de Investigación Nebrija en Cognición, Universidad Antonio de Nebrija, Madrid, Spain

[c] Department of Culture, Communication and Media, IOE, UCL's Faculty of Education and Society, University College London, UK

Declaration of Conflicting Interests: None.

*Correspondence concerning this article should be addressed to Irini Mavrou, Departamento de Lenguas Aplicadas, Facultad de Lenguas y Educación, Universidad Antonio de Nebrija, Santa Cruz de Marcenado 27, 28015, Madrid, Spain.

Moral Judgements in a Foreign Language: Expressing Emotions and Justifying Decisions

Aim: Previous evidence suggests that language influences bilinguals' moral judgements. One explanation for this phenomenon is that using a second language (L2) attenuates emotional arousal, thus leading to more rational decisions. This study examined whether bilinguals' moral arguments and emotional vocabulary are influenced by the language – first language (L1) or L2 – in which a moral dilemma is presented.

Methodology and Data Analysis: A mixed-methods design was employed. We analysed the emotional vocabulary used by 204 Spanish-English bilinguals when making moral judgements and expressing their emotions in response to a highly emotional moral dilemma, as well as the type of arguments they employed to justify their moral decisions in L1 and L2.

Findings: The participants were more emotional in their L1, as reflected in the arguments they used to justify their decisions. This finding was supported by a significantly lower number of emotional words in the L2. Moreover, the effect of language on moral judgements was mediated by the participants' emotions.

Originality: This study is the first to qualitatively examine the types of arguments underlying bilinguals' moral decision-making in their L1 and in their L2. Moreover, the analysis of verbal emotional expressions in relation to moral decisions adds to the findings of previous research that was based almost exclusively on forced-choice measures, and further supports the hypothesis that the reduction of emotional arousal in an L2 modulates individuals' moral judgements.

Implications: The results have implications for L2 teaching and pedagogy. The L2 curriculum should include instruction in emotional vocabulary and should engage learners in discussions that require argumentation and critical thinking about strong emotional content. This may assist bilinguals not only to express their internal affective states more efficiently, but also experience the intensity of L2 emotionally charged words in a similar way as they do in their L1.

Introduction

Moral judgement and moral reasoning are concomitant with human affective reactions (Blanchette & Richards, 2010; Greene et al., 2001; Haidt, 2001; Loewenstein & Lerner, 2003; Wheatley & Haidt, 2005) and are essential in order to build a coherent representation of the world around us (Blanchette & Richards, 2010). We live in a globalised world; therefore, an increasing number of people are confronted with situations in which they need to express themselves and to make decisions in a language other than their first language (L1). As a result, over the last decade, several studies have focused on the influence of language on cognitive processes, such as moral decision making. These studies have suggested that people often make different moral decisions based on the language – L1 or a second language (L2) – in which the moral dilemma is presented (e.g. Costa, Foucart, Hayakawa, et al., 2014; Geipel et al., 2015a, 2015b, 2016; Hadjichristidis et al., 2019; Hayakawa et al., 2016; but see Muda et al., 2020; Wong & Ng, 2018). In fact, it has been observed that an L2 that has been acquired after puberty usually elicits weaker emotional responses compared to the L1 (Dewaele, 2010; Harris et al., 2003; Pavlenko, 2012). This may lead to a reduction in the emotionality of moral dilemmas and thus affect moral decision making (Costa et al., 2017). However, there is still scarce evidence to support this claim, and studies that have addressed this issue have almost exclusively employed forced-choice tasks to assess bilinguals' emotions arising from moral dilemmas (Driver, 2022; Geipel et al., 2015b). In this study, we examined whether the language – L1 versus L2 – in which a moral dilemma was presented influenced the emotional vocabulary and the arguments used by Spanish-English bilinguals to justify their moral decisions and to express emotions as a result of moral decision making in both the L1 and the L2. In addition, we

sought to explore whether the effect of language on moral judgements was mediated by emotions, as expressed verbally by the bilinguals themselves.

Moral Foreign Language Effect

Until recently, language had not been considered as a factor that could play an important role in moral decision making, but rather as a vehicle with the mere function of transmitting decisions (Hayakawa et al., 2016). In fact, variations in moral judgements have usually been attributed to contextual factors such as the time of the day and previous experience or familiarity with a situation (Costa et al., 2019). However, the role of language in moral decision making has recently gained attention, with an increasing number of studies positing that the language we use affects our moral judgements (e.g. Brouwer, 2019; Costa, Foucart, Hayakawa, et al., 2014; Dylman & Champoux-Larsson, 2020; Geipel et al., 2015a, 2015b, 2016; Hayakawa et al., 2017).

Costa, Foucart, Hayakawa, et al. (2014) were the first to observe the influence of the L2 on moral judgement. They reported that the use of an L2 increased utilitarian (as opposed to deontological) responses. Deontological principles lead to actions that are considered ethical even if they do not achieve the greatest possible good, while utilitarian decisions could lead to physical harm but are associated with the most positive outcomes (Greene, 2008). Costa, Foucart, Hayakawa, et al. (2014) presented native speakers of various languages who had learned an L2 in instructional settings with the classic footbridge dilemma and the switch dilemma (Foot, 1967; Thomson, 1985). In both dilemmas, a runaway trolley has broken down and cannot be stopped, and five people who are on the track will be killed if no action is taken. In the switch dilemma, the train can be diverted to another track on which a man is working. In the footbridge dilemma, one can opt to push a man from a bridge onto the track to stop the

train. In both versions, the participants need to decide whether to kill one person to save five. The results revealed that, in the footbridge dilemma, the participants were twice as willing to push the man from the bridge onto the track (utilitarian option) when the dilemma was presented in their L2 than when it was presented in their L1. However, no differences were observed in the switch dilemma, in which around 80% of the participants decided to divert the train regardless of the language in which the dilemma was presented. The difference in the results regarding the two dilemmas can be explained by the level of emotionality, which is higher in the footbridge version (Greene et al., 2009). In light of these findings, the authors argued that people tend to be more utilitarian when they make moral decisions in their L2 because of the emotional distance when thinking in the L2 as opposed to in the L1. These findings have been replicated in subsequent studies (e.g. Brouwer, 2021; Cipolletti et al., 2016; Geipel et al., 2015b; Hayakawa et al., 2017). This influence of language on moral decision making is known as the *Moral Foreign Language effec*t (MFLe).

The mechanisms underlying the (M)FLe are still not well understood and remain equivocal. At least three non-mutually exclusive explanations have been proposed. The first relies on the claim that using an L2 involves a higher cognitive load, which was originally thought to favour the deliberative processing associated with rational decisions (Costa, Foucart, Arnon, et al., 2014; Costa, Foucart, Hayakawa, et al., 2014; Keysar et al., 2012); however, it was later demonstrated that it decreased deontological (Hayakawa et al., 2017) or both types of processing (Hennig & Hütter, 2021; Muda et al., 2018). A second explanation for the MFLe is that of an increased psychological distance (i.e. the ability to separate ourselves from other instances, such as an individual or an event) when using an L2 (Corey et al., 2017; Hayakawa et al., 2016). This might be caused by different mental representations at the level of construal during moral

decision making in the L1 versus the L2 (Shin & Kim, 2017). However, the influential role of psychological distance in decision making and, particularly in moral choices, is still debated (Eyal et al., 2008; Gong & Medin, 2012; Žeželj & Jokić, 2014). The third explanation refers to the reduction of emotional arousal that certain types of moral dilemmas may evoke (McFarlane & Cipolletti Perez, 2020). As mentioned previously, this study aimed to investigate this hypothesis, which we refer to here as the *Reduced Emotionality Hypothesis* (REH) and present in more detail in the next section.

The Reduced Emotionality Hypothesis

According to the REH, processing highly emotionally salient moral dilemmas in the L2 may elicit weaker emotional reactions than in the L1 (Costa, Foucart, Hayakawa, et al. 2014; Hayakawa et al., 2017). Greene (2008) pointed out that deontological judgements are usually related to moral dilemmas that evoke high levels of emotional arousal; therefore, a reduction in emotional reactivity would lead to more utilitarian decisions. In other words, people are more eager to take an action for the greater good when they are less biased by their emotions. Following this reasoning, since affective processing is reduced in the L2 – which is usually acquired later in life and in classroom settings –, compared to the L1 that is acquired in naturalistic environments (Pavlenko, 2012), more utilitarian responses would be expected when moral dilemmas are presented in the L2. This is indeed what has been observed.

The reduction of emotionality in the L2 has already been reported in the processing of emotional stimuli, such as taboo words or expressions of love, which trigger lower emotional arousal in the L2 than in the L1 (Caldwell-Harris & Ayçiçeği-Dinn, 2009; Dewaele, 2004, 2008; Harris et al., 2003). However, there is still little evidence to support the impact of emotional arousal on the MFLe. For example, Geipel

et al. (2015b) presented 161 late Chinese-English bilinguals (Experiment 2) with high (i.e. the footbridge) and low (i.e. the switch) emotionally salient moral dilemmas. Participants were asked to judge the appropriateness of the action described in each dilemma and to rate the extent to which each dilemma elicited specific emotions, such as being upset, worried, or sad. The authors found that, although language appeared to play a role in moral judgement in the footbridge dilemma, emotional attenuation did not appear to mediate the effect of the L2 on moral judgements. The authors argued that people may not be less emotional when they are confronted with a moral scenario in the L2, but that some moral transgressions are likely to be perceived as less condemnable in the L2 than they are in the L1. In fact, most moral transgressions are based on norms that have been learned – directly or indirectly – during childhood through social interactions in naturalistic contexts in which the L1 is used; therefore, these norms are believed to have more impact in L1 contexts.

In a more recent study by Caldwell-Harris and Ayçiçeği-Dinn (2020), the REH was tested using physiological techniques. Specifically, 51 bilingual participants (Turkish-English) evaluated six moral dilemmas concerning non-ethical (selfish) or ethical actions in the L1 or in the L2, and skin conductance responses were measured. Of interest, the ethical agreement ratings were higher in the L1. Furthermore, although the participants' skin conductance responses were more pronounced when they were presented with ethical statements in both language conditions, the difference in the magnitude of the skin conductance responses between non-ethical and ethical statements were stronger in the L1, while these differences were minor in L2 and were attributed to the reduction of emotional arousal when individuals encountered an unethical action in an L2 context.

Overall, previous findings and the types of measures used do not allow us to determine whether the use of an L2 reduces emotional responses related to emotionally charged moral dilemmas. The current study extends the previous work by inviting participants to freely express their arguments and their emotions underlying a moral decision in their L1 and in their L2. Methodologically speaking, the study differs from most previous work in that it used mediation analysis as a more direct way of investigating whether the MFLe is mediated by weakened affective responses to a highly emotional moral dilemma (Costa, Foucart, Hayakawa, et al., 2014). In accordance with the REH, it was hypothesised that participants would use fewer emotional arguments and fewer emotional words (high-arousal words) to justify their decisions and to express their emotions in the L2. We also predicted that the MFLe would be mediated by decreased emotionality in the L2. Before reporting on the current study, we will briefly describe what emotional expression involves.

Emotional Expression

As mentioned previously, the role of emotions in moral judgements has been emphasised using different types of moral dilemmas (Greene et al., 2004; Greene et al., 2001; Szekely & Miu, 2014, amongst others). Highly emotional moral dilemmas tend to provoke stronger emotional responses (Cecchetto et al., 2017) and lead to more deontological inclinations (Greene, 2007). As McFarlane and Cipolletti Perez (2020) pointed out, "deontological responses are always perfectly correlated with emotional arousal and utilitarian responses are always correlated with a lack of this arousal in a range of moral dilemmas" (p. 4). The question that arises is how an emotional response is manifested verbally – beyond any yes or no responses – when people are asked to justify their moral decisions in their L1 or their L2. We consider that an analysis of

emotional expressions could shed light on the hypothesis that processing information in the L2 leads people to think less emotionally, and thus to make more utilitarian judgements when encountering moral dilemmas with elevated emotional salience.

Caffi and Janney (1994) defined emotional expression as a spontaneous, involuntary and explosive act of emotion that is part of speech. According to Pavlenko (2008), emotion words refer directly to an emotional state or process and are used to express the emotional state or process of something or someone, or to express what one feels. Emotion-related words describe behaviours associated with emotions without emphasising the emotion itself. With regard to emotionally charged words, their processing can express or generate emotions without referring explicitly to an emotion.

Several authors who are interested in the analysis of emotional vocabulary in linguistic corpora consider that emotional words can be described in terms of arousal and valence (Bradley & Lang, 1999; Stadthagen-Gonzalez et al., 2017; Warriner et al., 2013, amongst others). Arousal is understood as the level of intensity (low or high) caused by a stimulus, while valence indicates how pleasant or unpleasant (positive, negative or neutral) a stimulus is perceived as being (Stadthagen-Gonzalez et al., 2017). Emotional words can be classified as positive or negative and as high or low in arousal. By contrast, neutral words usually have intermediate levels of arousal and are perceived as words without emotional connotation.

An increasing number of recent studies have examined the affective properties (valence and arousal) of the emotional vocabulary used by L2 speakers (e.g. Jiménez Catalán & Dewaele, 2017; Kyriakou & Mavrou, in press; Mavrou, 2021; Mavrou & Bustos-López, 2018; Pavlenko & Driagina, 2007; Pérez-García & Sánchez, 2020; Vidal Noguera et al., 2022). For example, Pavlenko and Driagina (2007) analysed and compared the size and richness of the emotional vocabulary in oral narratives produced

by monolingual Russian speakers, monolingual American speakers, and Americans who were highly proficient in Russian. Their findings showed that American learners of Russian produced a significantly lower proportion of emotion word. Mavrou and Bustos-López (2018) focused on the valence of the emotional vocabulary used by migrant learners of Spanish in the speaking section of a certification exam. Around one third of the discourse of these migrants included emotional words, particularly positively valenced words. The proficiency level in the target language was positively linked to the number of emotional words, while longer stays in the host country led to a greater use of negative words. In a more recent study, Kyriakou and Mavrou (in press) explored the emotional vocabulary that 62 English-Spanish bilinguals used to justify their moral decisions and to express their emotions after reading an everyday moral dilemma – the cheater's dilemma – in their L1 or L2. They found that bilinguals used a greater and more varied emotional vocabulary (positive, negative and high-arousal words) in their L1. A content analysis of this vocabulary further revealed that the predominant emotions in the L1 were fear, depression, disappointment, hurt and remorse, while the participants experienced the same negative emotions, albeit to a lesser extent, in the L2. Using a similar methodological approach, the present study compared the number of high-arousal words used by bilingual speakers in their L1 and L2 to justify their moral decisions and to express their emotions as elicited by a highly emotional moral dilemma. Following the REH, we hypothesised that participants would use a greater number of high-arousal words in their L1 than in their L2.

The Current Study

The purpose of this study was to examine the emotional vocabulary and the types of arguments underling moral judgements when Spanish-English bilinguals were

presented with the footbridge dilemma in their L1 (Spanish) and in their L2 (English).

To accomplish this, the number of high-arousal words that the participants used in their

arguments was calculated. In addition, the participants' arguments were analysed

qualitatively in order to elucidate whether the language used (L1 versus L2) modified

the way in which they behaved and reasoned in a dilemmatic situation, and particularly

the extent to which these arguments were driven by emotions or deliberative thinking.

Furthermore, we examined whether the participants' affective responses

(operationalised as the number of high-arousal words) mediated the effect of language

on moral judgements. The study was approved by the Research Ethics Committee of

Nebrija University (Reference numbers: UNNE-2020-006, UNNE-2021-001) and

followed the principles expressed in the Declaration of Helsinki.

Method

Participants

Two hundred and thirty-four participants, who were recruited via different social

media platforms, took part in the study. To ensure that the participants had at least an

intermediate level of English that would allow them to understand the dilemma, their

proficiency level was assessed using the Lexical Test for Advanced Learners of English

(LexTALE) (Lemhöfer & Broersma, 2012). In addition, the participants completed a

language background questionnaire asking them to report their self-perceived

proficiency levels in English in writing, speaking, listening and reading abilities using a

7-point Likert scale, as well as the percentage of daily use of English with their family,

friends, and when reading and watching television. Only data from the participants who

were native speakers of Spanish, spoke English as an L2, and had not been raised

speaking English at home (Hayakawa et al., 2017) were included in the analyses. The

participants who did not answer all the questions and those who indicated that they were aware of the footbridge dilemma were excluded ($n = 30$). Thus, the final sample consisted of 204 participants; 102 were assigned to the L1 condition and the remaining to the L2 condition. Participation was voluntary and the respondents did not receive compensation. We obtained the participants' informed consent using a written consent protocol.

Participants' details are presented in Table 1. The participants in the L2 condition slightly outperformed those in the L1 condition in terms of overall exposure to English, self-perceived overall proficiency, writing and speaking skills in English, and LexTALE scores; however, the average LexTALE scores in both language conditions were equivalent to an upper-intermediate B2 level according to the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001). It is quite unlikely that these differences had an influence on the results since English was not a requirement for responding to the dilemma in the L1 condition. On the other hand, the scores obtained by the participants in the L2 condition proved that their proficiency level in English was sufficient to understand the English version of the dilemma and to justify their moral decisions in English. It is also important to note that the participants in the two language conditions were matched for age and gender, which are factors that have been found to affect decision making (e.g. Armstrong et al., 2019; Fumagalli et al., 2010; McNair et al., 2018).

Table 1
Descriptive statistics for the two experimental groups

	L1 M (SD)	L2 M (SD)	t	p	d
Participants (number)	102	102			
Females (number)[†]	37	31			
Age (in years)	34.76 (6.58)	33.51 (8.52)	1.177	.241	0.165
Age of acquisition of English (in years)	9.31 (5.57)	8.47 (4.33)	1.207	.229	0.169
Number of participants who had lived in an English-speaking country	41	35			
Months spent in an English-speaking country	18.54 (19.70)	17.86 (33.21)	0.110	.912	0.025
Exposure to English (% of time)[††]	31.47 (18.99)	36.96 (18.98)	-2.065	.040[*]	-0.289

Self-perceived proficiency in English[†††]	4.64 (1.23)	4.97 (0.99)	-2.162	.032[*]	-0.303
Self-perceived reading ability in English	4.90 (1.23)	5.21 (1.10)	-1.858	.065	-0.260
Self-perceived writing ability in English	4.33 (1.29)	4.69 (1.11)	-2.094	.037[*]	-0.293
Self-perceived speaking ability in English	4.43 (1.37)	4.80 (1.27)	-2.018	.045[*]	-0.283
Self-perceived listening ability in English	4.87 (1.38)	5.20 (1.22)	-1.777	.077	-0.249
LexTALE	67.83 (11.23)	71.46 (11.53)	-2.276	.024[*]	-0.319

Note. [*]$p \leq .05$. [†] One participant did not provide information regarding gender. [††] Exposure to English at work, on the Internet, with family and friends, while watching movies and reading books, and other. [†††] The values range from 1 (= very poor) to 7 (= native-like).

Materials and Procedure

The materials were distributed and completed online via the QuestionPro survey platform (Bhaskaran, 2002). After providing their consent, the participants were presented with the footbridge dilemma (Appendix A) either in their L1 Spanish (the same text as used in Costa, Foucart, Hayakawa, et al., 2014) or in their L2 English, and were asked to indicate whether they would push the man onto the track using a binary (yes/no) scale. A more sensitive 7-point Likert scale was then used to collect the participants' judgements about the moral permissibility of their decisions (1 = totally impermissible, 4 = permissible, 7 = totally permissible) (Geipel et al., 2015a). Lower values indicated more deontological responses, while higher values indicated more utilitarian responses. Although no time limit was set because reading in an L2 usually takes longer than it does in the L1, the participants were requested to answer as quickly as possible to avoid long periods of reflection. The participants were also informed that there were no correct or incorrect answers. After indicating their decisions, they were invited to respond to open-ended questions related to the dilemma. Specifically, they were asked to explain their decisions and justify them, and to describe in detail how they felt while they were reading the dilemma and during (or after) making their decisions, as well as the emotions their decisions evoked and why (Appendix B). Lastly, the participants took the LexTALE and completed a language background questionnaire.

Emotional Vocabulary Analysis

To examine whether language influenced the emotional vocabulary used to justify moral decisions, we analysed the degree of arousal of the emotional words employed by the participants to justify their moral judgements and to express the emotions triggered by the moral dilemma and their moral choices. The participants' answers were transcribed first, and both content and function words were lemmatised. As emotional words are always content words, we excluded grammatical words from the analysis and only calculated the number of lexical tokens and types (Pavlenko & Driagina, 2007) using the programme V_Words v2.0 (Meara & Miralpeix, 2016). Lexical types were later analysed in order to identify all the high-arousal words in Spanish and English. To ensure that any differences in the number of high-arousal words across language conditions were not due to limited general vocabulary in L2, we also analysed and compared the number of low-arousal words in the L1 and in the L2. Different normative databases were used to accomplish this. The affective ratings for emotional words in Spanish were obtained from three normative databases (Hinojosa et al., 2016; Redondo et al., 2007[1]; Stadthagen-Gonzalez et al., 2017[2]) included in the online web-based search engine emoFinder (Fraga et al., 2018). For English, the Affective Norms for English Words (ANEW; Bradley & Lang, 1999) and Warriner et al.'s (2013) set of norms (which is a complement to ANEW) were used. The words included in these databases are rated using a 9-point scale. Following Guasch et al. (2016), words with an average score between 1.50 and 5.41 were grouped as low-arousal words, whereas words with scores ranging from 5.43 to 8.40 were classified as high-arousal words. When computing the total number of high- and low-arousal words for each participant, multiple repetitions of the same word(s) were taken into account; for example, if a

participant used the same high-arousal word three times, this word was counted as three tokens.

Qualitative Analysis of Arguments to Justify Moral Decisions

Inductive content analysis techniques (Elo & Kyngäs, 2008) were used in order to examine whether the language condition (L1 versus L2) influenced the types of arguments that the participants used to justify their moral judgements. A content analysis allows researchers to identify similar patterns and to group them according to the same categories or themes (Graneheim et al., 2017). This technique entails the three phases of preparation, organising and reporting (Elo & Kyngäs, 2008). In the preparation phase, the obtained data were transcribed and read several times in order the researchers to familiarise themselves with the data and to identify all the sentences that contained moral arguments. A moral argument was conceptualised as a verbal activity or attempt made by the participants to justify their moral decisions. Therefore, only sentences involving moral arguments were taken into account. In the organising phase, open coding was implemented using notes and highlights. Notes were taken while reading the participants' answers, while moral arguments were highlighted. We then classified arguments with similar content under higher-order headings. In the final phase, we formulated types of arguments based on the entire set of moral arguments provided by the participants, and some types of arguments that had similar content were grouped together in order to reduce the number of arguments and to facilitate the data interpretation. We divided these types of arguments into two principal categories, namely deontological (emotional) and utilitarian (rational). Finally, we calculated the frequency with which each moral argument appeared in the corpus.

<center>**Results**</center>

Moral judgement (yes/no)

We first examined whether the MFLe reported in previous studies could be replicated (Costa, Foucart, Hayakawa, et al., 2014; Hayakawa et al., 2017, amongst others) by analysing the yes/no responses to the dilemma. To accomplish this, we used both NHST procedures and Bayesian analysis. In the L1 condition, 8.82% of the participants decided to push the man (utilitarian decision), while 32.35% did so in the L2 condition ($\chi^2(1,204) = 17.270$, $p < .001$, $\varphi = .291$, $BF_{10} = 1000.736$). The odds were 4.94 times greater that individuals would choose the deontological option in their L1 (95% CI [2.22, 11], $Z = 3.914$, $p < .001$). These results replicated the MFLe.

Moral permissibility

With regard to the moral permissibility of the decision to push the man, the mean ratings were compared using independent samples *t*-tests, as the assumption of equality of variances was met ($F = 0.902$, $p = .343$). Cohen's *d* and Bayes factors were also calculated. The participants in the L1 condition rated the decision to push the man as being less permissible ($M = 2.17$, $SD = 1.59$) than did the participants in the L2 condition ($M = 2.76$, $SD = 1.72$), and these differences were statistically significant $t = -2.578$, $p = .011$, Cohen's $d = -0.361$).

Emotional vocabulary

Descriptive statistics for high and low arousal words by language condition (Spanish L1 and English L2) are summarised in Table 2. Comparison of means tests for the number of high- and low-arousal words were conducted using both *t*-tests and the Welch's *t*-test because the assumption of equal variances was not met. Cohen's *d* and

Bayes factors were also calculated (see Table 3). The results showed that the number of high-arousal words was considerably higher in the L1 ($d = 1.66$ representing a large effect size), while the number of low-arousal words was significantly higher in the L2 ($d = 0.91$).

Table 2
Descriptive statistics for the number of high and low arousal words in L1 and L2

Condition	Spanish L1				English L2			
	M	SD	Skp	Ku	M	SD	Skp	Ku
High arousal words	9.67	5.01	0.88	0.40	3.19	2.32	1.13	0.99
Low arousal words	8.31	5.92	1.19	1.62	16.13	10.65	1.44	2.05

Table 3
Differences in the number of high and low arousal words between L1 and L2

	Test	Statistic	Df	p	Cohen's d	B_{10}
High arousal words	Student	11.852	202	< .001	1.660	4.598e +21
	Welch	11.852	142.546	< .001	1.660	
Low arousal words	Student	-6.477	202	< .001	-0.907	1.270e +7
	Welch	-6.477	158.024	< .001	-0.907	

Mediation analysis

A bootstrapping method was also conducted using SPSS Process Macro version 4.1 (Hayes, 2022) to explore whether the participants' emotions (operationalised as the number of high-arousal words) mediated the effect of language on moral judegment (deontological versus utilitarian decisions). The results of the regression analysis revealed that language was a significant predictor of the number of high-arousal words (b = -6.48, t = -11.85, p < .001, R^2 = .41, Bootstrap 95% CI [-7.55, -5.40]). While controlling for the number of high-arousal words (mediator), the results of the second regression model showed that language was not a significant predictor of moral judgements (b = 0.71, Z = 0.13, p = .89, Bootstrap 95% CI [-0.97, 1.11]). The results of the indirect effect based on 5000 bootstrap samples led to a significant indirect positive relationship between the language condition and moral judgement mediated by the participants' emotions (b = 2.21, Bootstrap 95% CI [1.01, 4.78]; see also Figure 1).
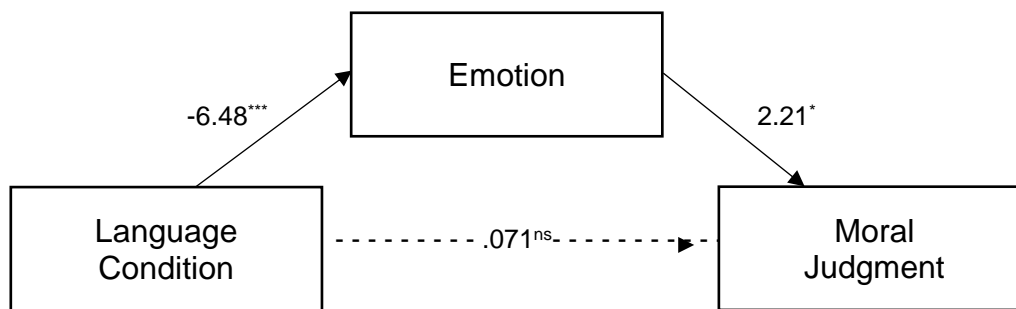
Figure 1. Illustration of the indirect effect of language on moral judgment. Number refers to unstandardised beta coefficients. $^*p < .05$, $^{***}p < .001$.

Moral arguments

A qualitative analysis of the arguments used by the participants to justify their moral decisions (see Appendices C and D for L2 English and L1 Spanish, respectively) revealed that the use of the L2 increased the number of utilitarian arguments dramatically ($n = 101$ compared to $n = 16$ in L1). Twenty-one different types of moral arguments related to the utilitarian decision were identified in the L2 compared to only 11 in the L1. In the L2 condition, 22.57% of the participants indicated that they made a rational decision because one death is better than five. Some of the participants further argued that it was a moral duty to defend the common good and that it was better to act than do nothing (for example, participant ID82 stated: "I feel terrible because I have just become a killer. I could have chosen to do nothing, but I'd be a passive killer in that case"). However, only 3.92% of the participants in the L1 condition indicated that five lives were worth more than one life. It is also important to mention that some of the participants in the L2 condition indicated that they hesitated before making their decisions and had contradictory feelings (8.82%), but they finally acted in favour of the common good. In addition, some of the participants explicitly stated that emotions had no place in their moral judgements (4.9%) or that they felt more relieved when choosing to save five lives instead of one (4.9%). These arguments support the claim that bilinguals are less emotionally affected when encountering an emotionally charged moral dilemma in their L2.

The participants who opted for the deontological choice in the L1 condition explained their arguments in a variety of ways. Two hundred and sixty-eight moral arguments were classified according to 37 categories representing deontological decisions as opposed to only 192 moral arguments classified according to 25 categories in the L2 condition. However, it is important to acknowledge the similarity of some of the moral arguments in both conditions. For example, the main deontological argument was that every life is precious and that no one has the right to decide about other people's lives. Although several participants (34.31%) in the L1 condition indicated that it was morally indefensible to kill one person regardless of the reason for doing so and that they would have felt guilty if they had chosen to kill the man, only 12.74% of the participants in the L2 condition reported the same argument.

In addition, 14 participants in the L1 condition stated that pushing the man onto the track would make them murderers and that this would weigh heavily on their consciences (for example, "*Soy incapaz de cargar con la muerte de alguien en mi conciencia* [I am not capable of killing someone because it will be a burden on my mind]", ID26). However, some of the participants (4.9%) in the L2 condition reported that they would be capable of killing the man if they knew any of the five people or if the man was a murderer. Others indicated that it was better not to intervene because accidents happen every day and the footbridge dilemma is not "your fault" (7.84%) ("It's way easier not to interfere, being a simple spectator and not actively killing someone. It's almost a natural response", ID11). Participants in the L1 condition also appeared to be slightly more concerned about the legal consequences of killing someone. As participant ID50 mentioned, "*Legalmente, si no me equivoco, tu acto no tiene legitimidad, por lo que a los ojos de la justicia eres un asesino* [Legally speaking,

if I am not mistaken, your act has no legitimacy, so in the eyes of the law you are a murderer]".

Discussion

The present study examined whether the language in which a highly emotional moral dilemma was presented (L1 versus L2) influenced bilinguals' moral arguments and emotional expressions. The results indicated that the participants used a significantly lower number of high-arousal words, as well as a greater number of low-arousal words, in their L2. These findings are consistent with evidence from cross-linguistic studies showing that bilinguals are less emotionally expressive in their L2 than in their L1 (Dewaele & Pavlenko, 2002; Pavlenko & Driagina, 2007). The higher number of low-arousal words in the L2 condition suggests that our participants had acquired and were able to use a wide range of vocabulary to express themselves in their L2. This result supports our hypothesis that the differences in the number of high-arousal words and emotional arguments across language conditions were rather driven by a reduced sensitivity to emotional vocabulary rather than by limited L2 vocabulary. As Altarriba (2003) argued, the representations of L1 emotion words in the memory are stronger because individuals acquire and use these types of words in a large variety of contexts, whereas L2 emotion words are mainly acquired in instructional settings, are applied less frequently in real-life situations, and thus do not activate the same associations as do L1 emotion words. For example, in Pavlenko and Driagina's (2007) study, difficulties in retrieving L2 emotional vocabulary led bilinguals to apply strategies such as paraphrasing and lexical borrowing in order to communicate their feelings and emotions in the L2.

However, it is also possible that people are less driven by their emotions and instincts when they make moral decisions in their L2 (Costa, Foucart, Hayakawa, et al., 2014), which in turn may lead to the use of fewer emotionally charged words when they are asked to express their emotions in the L2. This second explanation is fully supported by the results of the mediation analysis and corroborates previous findings suggesting a reduction in the emotional reactivity of bilingual speakers when they use certain types of emotion words and expressions in their L2, such as taboo words and reprimands (e.g. Caldwell-Harris & Ayçiçeği-Dinn, 2009; Dewaele, 2004, 2008; Harris et al., 2003). However, our findings also contradict evidence from the field of moral decision making that failed to demonstrate the influence of emotional blunting in the L2 on bilinguals' moral judgements (Driver, 2022; Geipel et al., 2015b). The reason for this discrepancy may be due to the methodology that was used to assess emotions in previous studies, which was mainly based on emotional arousal rating scales and lists of predetermined basic emotions that do not fully capture how emotions exert influence on moral judgements and moral reasoning. In this regard, the present study introduced an alternative way to examine the REH via an in-depth analysis of the emotional vocabulary that bilinguals used to verbalise their emotions and feelings in response to a moral dilemma.

With regard to the arguments used to justify moral decisions, the main finding was that the participants used a considerably higher number of emotional arguments in their L1 than they did in their L2. Although some types of arguments were similar in both conditions, the participants in the L1 condition expressed more categorical opinions and, on many occasions, appeared to be more empathetic towards the overweight man, or emphasised feelings of guilt (see Cohen et al., 2012, and Huebner et al., 2009, for similar evidence), an emotion that was less pronounced in the L2. The participants in

the L1 condition were also more concerned about the legal consequences of killing a person and emphasised the importance of not being a murderer regardless of the circumstances more strongly. Studies by Geipel et al. (2015b) and Białek et al. (2019) concluded that moral norm violations were judged more harshly when a moral dilemma was presented in the L1, while people tended to pay less attention to moral norms in the L2; that is, they were less concerned about the consequences of their utilitarian decisions and norm transgressions.

This analysis also revealed that the use of the L2 prompted a greater number of utilitarian moral arguments. Participants in the L2 condition tended to choose to kill someone to save more lives based on arguments such as "the end justifies the means", "it is better to act than do nothing" or "I don't know any of them, so I chose to save more people". Some of them also admitted that their decisions were instinctive, were made without feelings and that they would feel more relieved knowing that they had saved five lives instead of one. The latter argument contradicts those provided by several participants in the L1 condition. Therefore, the greater use of utilitarian arguments in the L2 provides further evidence for the assumption that individuals react less emotionally to affective stimuli in languages other than their L1 (Caldwell-Harris, 2014; Costa, Foucart, Hayakawa, et al., 2014; Dewaele, 2004; Iacozza et al., 2017).

Nonetheless, it is interesting to note that some participants argued that they were unsure or not confident about their decisions and had contradictory feelings. These findings are in line with previous research indicating that participants were less confident when they made moral judgements in their L2 (Geipel et al., 2015a). As deliberative thinking gives rise to more confidence during the decision-making process (Mata et al., 2013), our findings appear to be consistent with the claim that "people are

more utilitarian when using a foreign language not because they think more, but because they feel less" (Hayakawa et al., 2017, p. 1396).

Limitations and Future Research

The present study has certain limitations that need to be acknowledged and taken into account in future work. Firstly, only one emotionally charged moral dilemma was used. Future studies should include moral dilemmas that differ in their emotionality in order to examine the effect of emotion on bilinguals' moral judgements in more depth and to determine whether emotionally charged moral dilemmas evoke more emotional responses in the L1 and in the L2 (see Horne & Powell, 2013, 2016). For example, it would be desirable to compare bilinguals' emotional vocabulary in their L1 and L2 using both high- and low-emotion dilemmas. Differences in emotional vocabulary for high – but not for low – emotion moral dilemmas would further support the REH. Secondly, the present study focused on a classic moral dilemma (the footbridge dilemma) that describes a conflict situation that it is unlikely to occur in real life. Therefore, future research should investigate the MFLe using more realistic dilemmas that bilinguals would likely encounter in their daily lives. A third limitation is related to the methodology used to assess emotional expression, which was based on the analysis of the emotional vocabulary and arguments used by the participants. These findings should be confirmed by future studies employing more direct measurements of emotional arousal to evaluate the viability of the REH, such as electrophysiological measures (see Greene et al., 2004; Harris et al., 2003, amongst others). Another limitation concerns participants' proficiency in English. Although all of them had an upper-intermediate B2 level (CEFR), the participants in the L2 condition slightly outperformed those in the L1 condition. Although it is unlikely that this had an

influence on the results, future laboratory studies should try to thoroughly control for differences in L2 proficiency level. Furthermore, we only analysed the MFLe with participants whose L1 was Spanish and who spoke English as an L2. More pairs of languages, particularly those that are underrepresented in the scientific literature, should be explored more thoroughly. In addition, the effect of modality has not been systematically investigated (Brouwer, 2019, 2021; Muda et al., 2020), and this study only analysed responses that were provided in the written mode. Considering that "oral narratives are more representative of spontaneous speech" (Pavlenko & Driagina, 2007, p. 218), future replication studies can attempt to identify the link between emotions and moral judgements via face-to-face, more naturalistic oral interactions. Moreover, it cannot be ruled out that individual differences in the participants' L2 reading proficiency may have influenced the results. For example, a recent meta-analysis (Stankovic et al., 2022) revealed that lower L2 reading proficiency was associated with more utilitarian judgements in emotionally charged moral dilemmas (but see Circi et al., 2021; Del Maschio et al., 2022). In order to advance research in the filed forward and to attain a better understanding of the MFLe, future studies should attempt to identify potentially relevant variables (such as emotional acculturation, linguistic distance between the L1 and the L2, and L2 reading abilities) and to explore how they contribute to bilinguals' moral choices.

Notes

[1] These authors carried out the Spanish adaptation of the ANEW.

[2] It is currently the largest Spanish database for arousal and valence dimensions.

References

Altarriba, J. (2003). Does cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism*, *7*(3), 305–322. https://doi.org/10.1177/13670069030070030501

Armstrong, J., Friesdorf, R., & Conway, P. (2019). Clarifying gender differences in moral dilemma judgments: The complementary roles of harm aversion and action aversion. *Social Psychological and Personality Science*, *10*(3), 353–363. https://doi.org/10.1177/1948550618755873

Bhaskaran, V. (2002). *QuestionPro* [platform]. Texas. https://www.questionpro.com

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*, Article 103855. https://doi.org/10.1016/j.jesp.2019.103855

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion*, *24*(4), 561–595. https://doi.org/10.1080/02699930903132496

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Technical report C-1). The Center for Research in Psychophysiology, University of Florida.

Brouwer, S. (2019). The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *Journal of Multilingual and Multicultural Development*, *40*(10), 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*(2), 223–230. https://doi.org/10.1017/S136672892000022X

Caffi, C., & Janney, R. W. (1994). Toward a pragmatics of emotive

  communication. *Journal of Pragmatics*, *22*(3-4), 325–373.

  https://doi.org/10.1016/0378-2166(94)90115-5

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign

  language: Theoretical implications. *Frontiers in Psychology, 5,* Article 1055.

  https://doi.org/10.3389/fpsyg.2014.01055

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native

  language. *International Journal of Psychophysiology, 71*(3)*,* 193–204.

  https://doi.org/10.1016/j.ijpsycho.2008.09.006

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2020). When using the native language

  leads to more ethical choices: Integrating ratings and electrodermal

  monitoring. *Language, Cognition and Neuroscience*, *36*(7), 885–901.

  https://doi.org/10.1080/23273798.2020.1818266

Cecchetto, C., Korb, S., Rumiati, R. I., & Aiello, M. (2017). Emotional reactions in

  moral decision-making are influenced by empathy and alexithymia. *Social*

  Neuroscience, 13(2), 226–240. https://doi.org/10.1080/17470919.2017.1288656

Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language

  effect. *Philosophical Psychology*, *29*(1), 23–40.

  https://doi.org/10.1080/09515089.2014.993063

Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on

  decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, *28*(4), 1131–

  1141. https://doi.org/10.3758/s13423-020-01871-z

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science, 21*(5), 355–359. https://doi.org/10.1177/0963721412454874

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*(7), 1109–1128. https://doi.org/10.1037/xlm0000356

Costa, A., Duñabeitia, J. A., & Keysar, B. (2019). Language context and decision-making: Challenges and advances. *Quarterly Journal of Experimental Psychology*, *72*(1), 1–2. https://doi.org/10.1177/1747021818789799

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition*, *130*(2), 236–254. https://doi.org/10.1016/j.cognition.2013.11.010

Costa, A., Foucart., A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PLoS ONE*, *9*(4), Article e94842. https://doi.org/10.1371/journal.pone.0094842

Costa, A., Vives, M. L., & Corey, J. D. (2017). On language processing shaping decision making. *Current Directions in Psychological Science*, *26*(2), 146–151. https://doi.org/10.1177%2F0963721416680263

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

Del Maschio, N., Crespi, F., Peressotti, F., Abutalebi, J., & Sulpizio, S. (2022). Decision-making depends on language: A meta-analysis of the Foreign Language Effect. *Bilingualism: Language and Cognition*, *25*(4), 617–630. https://doi.org/10.1017/S1366728921001012

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the
    speech of multilinguals. *Journal of Multilingual and Multicultural
    Development*, *25*(2-3), 204–222. https://doi.org/10.1080/01434630408666529

Dewaele, J.-M. (2008). The emotional weight of *I love you* in multilinguals'
    languages. *Journal of Pragmatics*, *40*(10), 1753–1780.
    https://doi.org/10.1016/j.pragma.2008.03.002

Dewaele, J.-M. (2010). "Christ fucking shit merde!" Language preferences for swearing
    among maximally proficient multilinguals. *Sociolinguistic Studies*, *4*(3), 595–614.
    https://doi.org/10.1558/sols.v4i3.595

Dewaele, J.-M., & Pavlenko, A. (2002). Emotion vocabulary in interlanguage.
    Language Learning, *52*(2), 263–322. https://doi.org/10.1111/0023-8333.00185

Driver, M. Y. (2022). Switching codes and shifting morals: How code-switching and
    emotion affect moral judgment. *International Journal of Bilingual Education and
    Bilingualism*, *25*(3), 905–921. https://doi.org/10.1080/13670050.2020.1730763

Dylman, A. S., & Champoux-Larsson, M. F. (2020). It's (not) all Greek to me:
    Boundaries of the foreign language effect. *Cognition*, *196*, Article 104148.
    https://doi.org/10.1016/j.cognition.2019.104148

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of
    Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-
    2648.2007.04569.x

Eyal, T., Liberman, N., & Trope, Y. (2008). Judging near and distant virtue and
    vice. *Journal of Experimental Social Psychology*, *44*(4), 1204–1209.
    https://doi.org/10.1016/j.jesp.2008.03.012

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford
    Review*, *5*, 5–15.

Fraga, I., Guasch, M., Haro, J., Padrón, I., & Ferré, P. (2018). EmoFinder: The meeting

point for Spanish emotional words. *Behavior Research Methods*, *50*(1), 84–93.

https://doi.org/10.3758/s13428-017-1006-3

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S.,

Lucchiari, C., Consonni D., Nordio, F., Pravettoni, G., Cappa, S., & Priori, A.

(2010). Gender-related differences in moral judgments. *Cognitive Processing*,

11(3), 219–226. https://doi.org/10.1007/s10339-009-0335-2

Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). How foreign language shapes

moral judgment. *Journal of Experimental Social Psychology*, *59*, 8–17.

https://doi.org/10.1016/j.jesp.2015.02.001

Geipel, J., Hadjichristidis C., & Surian L. (2015b). The foreign language effect on

moral judgment: The role of emotions and norms. *PLoS ONE, 10*(7), Article

e0131529. https://doi.org/10.1371/journal.pone.0131529

Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the

contribution of intentions and outcomes to moral judgment. *Cognition*, *154*, 34–

39. https://doi.org/10.1016/j.cognition.2016.05.010

Gong, H., & Medin, D. L. (2012). Construal levels and moral judgment: Some

complications. *Judgment and Decision Making*, *7*(5), 628–638.

Graneheim, U. H., Lindgren, B. M., & Lundman, B. (2017). Methodological challenges

in qualitative content analysis: A discussion paper. *Nurse Education Today*, *56*,

29–34. https://doi.org/10.1016/j.nedt.2017.06.002

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory

of moral judgment explains. *Trends in Cognitive* Sciences, *11*(8), 322–323.

https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–80). MIT Press.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364–371. https://doi.org/10.1016/j.cognition.2009.02.001

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, *48*(4), 1358–1369. https://doi.org/10.3758/s13428-015-0684-y

Hadjichristidis, C., Geipel, J., & Keysar, B. (2019). The influence of native language in shaping judgment and choice. *Progress in Brain Research*, *247*, 253–272. https://doi.org/10.1016/bs.pbr.2019.02.003

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295x.108.4.814

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first than in a second language. *Applied Psycholinguistics, 24*(4), 561–578. https://doi.org/10.1017/S0142716403000286

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language

changes our choices. *Trends in Cognitive Sciences*, *20*(11), 791–793.

https://doi.org/10.1016/j.tics.2016.08.004

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking

more or feeling less? Explaining the foreign-language effect on moral judgment.

Psychological Science, *28*(10), 1387–1397.

https://doi.org/10.1177/0956797617720944

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process

analysis: A regression-based approach* (3rd ed.). The Guilford Press.

Hennig, M., & Hütter, M. (2021). Consequences, norms, or willingness to interfere: A

proCNI model analysis of the foreign language effect in moral dilemma judgment.

Journal of Experimental Social Psychology, 95, Article 104148.

https://doi.org/10.1016/j. jesp.2021.104148

Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U.,

Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of

875 Spanish words for five discrete emotional categories and two emotional

dimensions. *Behavior Research Methods*, *48*(1), 272–284.

https://doi.org/10.3758/s13428-015-0572-5

Horne, Z., & Powell D. (2013). More than a feeling: When emotional reactions don't

predict moral judgments. In M. Knauf, M., Pauven, N. Sebanz, & I. Wachsmuth

(Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science

Society* (pp. 615-620). Cognitive Science Society.

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral

dilemmas? *PLoS ONE*, *11*(7), Article e0154780.

https://doi.org/10.1371/journal.pone.0154780

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6. https://doi.org/10.1016/j.tics.2008.09.006

Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PloS ONE*, *12*(10), Article e0186027. https://doi.org/10.1371/journal.pone.0186027

Jiménez Catalán, R. M., & Dewaele, J. M. (2017). Lexical availability of young Spanish EFL learners: Emotion words versus non-emotion words. *Language, Culture and Curriculum*, *30*(3), 283–299. https://doi.org/10.1080/07908318.2017.1327540

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, *23*(6), 661–668. https://doi.org/10.1177/0956797611432178

Kyriakou, A., & Mavrou, I. (in press). *¿Eres muy emocional*? I don't think so. How does language influence our emotional responses to everyday moral dilemmas? In A. Blanco Canales & S. Martín Leralta (Eds.), *Emotion and identity in foreign language learning*. Peter Lang.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 619–642). Oxford University Press.

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of*

Personality and Social Psychology, *105*(3), 353–373.

https://doi.apa.org/doi/10.1037/a0033640

Mavrou, I. (2021). Emotional intelligence, working memory, and emotional vocabulary

in L1 and L2: Interactions and dissociations. *Lingua*, *257*, Article 103083.

https://doi.org/10.1016/j.lingua.2021.103083

Mavrou, I., & Bustos-López, F. (2018). Vocabulario emocional en la producción oral en

español como lengua de migración. *Doblele. Revista de Lengua y Literatura*, *4*,

41–60. https://doi.org/10.5565/rev/doblele.40

McFarlane, S., & Cipolletti Perez, H. (2020). Some challenges for research on emotion

and moral judgment: The moral foreign-language effect as a case

study. *Diametros*, *17*(64), 56–71. https://doi.org/10.33392/diam.1476

McNair, S., Okan, Y., Hadjichristidis, C., & de Bruin, W. B. (2018). Age differences in

moral judgment: Older adults are more deontological than younger adults. *Journal

of Behavioral Decision Making*, *32*(1), 47–60. https://doi.org/10.1002/bdm.2086

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Multilingual

Matters.

Muda, R., Niszczota, P., Białek, M., & Conway, P. (2018). Reading dilemmas in a

foreign language reduces both deontological and utilitarian response

tendencies. *Journal of Experimental Psychology: Learning, Memory, and

Cognition*, *44*(2), 321–326. https://doi.org/10.1037/xlm0000447

Muda, R., Pieńkosz, D., Francis, K. B., & Białek, M. (2020). The moral foreign

language effect is stable across presentation modalities. *Quarterly Journal of

Experimental Psychology*, *73*(11), 1930–1938.

https://doi.org/10.1177/1747021820935072

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual

lexicon. *Bilingualism: Language and Cognition*, *11*(2), 147–164.

https://doi.org/10.1017/S1366728908003283

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied

cognition? *International Journal of Psychology*, *47*(6), 405–428.

https://doi.org/10.1080/00207594.2012.743665

Pavlenko, A., & Driagina, V. (2007). Russian emotion vocabulary in American

learners' narratives. *The Modern Language Journal*, *91*(2), 213–234.

https://doi.org/10.1111/j.1540-4781.2007.00541.x

Pérez-García, E., & Sánchez, M. J. (2020). Emotions as a linguistic category:

Perception and expression of emotions by Spanish EFL students. *Language,*

Culture and Curriculum, 33(3), 274–289.

https://doi.org/10.1080/07908318.2019.1630422

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of

ANEW (Affective Norms for English Words). *Behavior Research*

Methods, *39*(3), 600–605. https://doi.org/10.3758/BF03193031

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

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017).

Norms of valence and arousal for 14,031 Spanish words. *Behavior Research*

Methods, *49*(1), 111–123. https://doi.org/10.3758/s13428-015-0700-2

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*, Article 105082. https://doi.org/10.1016/j.bandl.2022.105082

Szekely, R. D., & Miu, A. C. (2014). Incidental emotions in moral dilemmas: The influence of emotion regulation. *Cognition & Emotion*, *29*(1), 64–75. https://doi.org/10.1080/02699931.2014.895300

Thomson, J. (1985). The trolley problem. *Yale Law Journal*, *94*(6), 1395–1415. https://doi.org/10.2307/796133

Vidal Noguera, C., Villar, C. M., & Mavrou, I. (2022). Seething with rage: Affective vocabulary of anger in the autobiographical memories of Spanish-German bilinguals. In I. Mavrou, M. Pérez Serrano, & J-.M. Dewaele (Eds.), *Recent advances in second language emotion research* (pp. 155-189). Thomson Reuters Aranzadi.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10),780–784. https://doi.org/10.1111%2Fj.1467-9280.2005.01614.x

Wong, G., & Ng, B. C. (2018). Moral judgement in early bilinguals: Language dominance influences responses to moral dilemmas. *Frontiers in Psychology*, *9*, Article 1070. https://doi.org/10.3389/fpsyg.2018.01070

Žeželj, I. L., & Jokić, B. R. (2014). Replication of experiments evaluating impact of psychological distance on moral judgment (Eyal, Liberman & Trope, 2008; Gong & Medin, 2012). *Social Psychology, 45*(3), 223–231. https://doi.org/10.1027/1864-9335/a000188