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

Andreas Kyriakou

Alice Foucart

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

Funding: This study is part of the project EMILIA (FFI2017-83166-C2-2-R) funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe” and was also supported by the Community of Madrid and the European Social Fund (IN.MIGRA3-CM, ref. H2019/HUM5772).

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.

Keywords
foreign language effect, moral decision making, emotional vocabulary, emotional arousal, bilingualism

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 effect (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

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

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 ≤ .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., 2007¹; Stadthagen-Gonzalez et al., 2017²) 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.
Results

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>
<thead>
<tr>
<th>Condition</th>
<th>Spanish L1</th>
<th>English L2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>High arousal words</td>
<td>9.67</td>
<td>5.01</td>
</tr>
<tr>
<td>Low arousal words</td>
<td>8.31</td>
<td>5.92</td>
</tr>
</tbody>
</table>

### 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 judgment (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).
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,

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 MFL in 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 field 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


https://doi.org/10.1177/0963721412454874


[https://doi.org/10.1017/S1366728908003283](https://doi.org/10.1017/S1366728908003283)

[https://doi.org/10.1080/00207594.2012.743665](https://doi.org/10.1080/00207594.2012.743665)

[https://doi.org/10.1111/j.1540-4781.2007.00541.x](https://doi.org/10.1111/j.1540-4781.2007.00541.x)

[https://doi.org/10.1080/07908318.2019.1630422](https://doi.org/10.1080/07908318.2019.1630422)


[https://doi.org/10.1007/s10936-017-9498-7](https://doi.org/10.1007/s10936-017-9498-7)


https://doi.org/10.1080/02699931.2014.895300

https://doi.org/10.2307/796133


https://doi.org/10.1111%2Fj.1467-9280.2005.01614.x


https://doi.org/10.1027/1864-9335/a000188