The auditory foreign-language effect of moral decision making in highly proficient bilinguals

Susanne Brouwer

To cite this article: Susanne Brouwer (2019): The auditory foreign-language effect of moral decision making in highly proficient bilinguals, Journal of Multilingual and Multicultural Development, DOI: 10.1080/01434632.2019.1585863

To link to this article: https://doi.org/10.1080/01434632.2019.1585863

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 01 Mar 2019.

Submit your article to this journal

Article views: 199

View Crossmark data
The auditory foreign-language effect of moral decision making in highly proficient bilinguals

Susanne Brouwer

Dutch Language and Culture Department, Radboud University, Nijmegen, the Netherlands

ABSTRACT

Previous research has shown that people make systematically different decisions when faced with a moral dilemma in a native than in a foreign language [e.g. Costa, A., A. Foucart, S. Hayakawa, M. Aparici, J. Apesteguia, J. Heafner, and B. Keysar. 2014. “Your Morals Depend on Language.” PLoS One 9 (4): e94842]. The aim of the current study is to test the limits of this so-called Foreign-Language Effect by examining (1) whether it holds for highly proficient bilinguals of a closely related language pair (i.e. Dutch-English), and (2) whether it can be replicated in an auditory setting. In Experiment 1, 60 Dutch-English bilinguals read moral dilemmas in Dutch or in English, whereas in Experiment 2, a different sample of 60 Dutch-English bilinguals listened to the same dilemmas. After reading or listening, participants’ task was to indicate whether the proposed action was appropriate or not. The results showed that the Foreign-Language Effect was absent in Experiment 1, but present in Experiment 2. These findings aid in understanding the robustness of the Foreign-Language Effect, revealing that in some contexts it may be overcome and/or inhibited, whereas in others it may be enhanced.

ARTICLE HISTORY

Received 6 September 2018
Accepted 18 February 2019

KEYWORDS

Moral decision making; foreign language; bilingualism; social and moral norms; foreign-language effect

Introduction

Individuals make thousands of decisions every day. Such decisions can be relatively small or may have a larger impact on life. For example, imagine you are walking in a shopping street and you find a wallet with lots of money. What will you do? Will you bring the wallet with the money to the police station or will you keep it? The ability to make moral decisions is an important aspect of social competence in humans. Previous research has shown that people make systematically different decisions when faced with a moral dilemma in a native (L1) than in a foreign language (L2) (e.g. Costa et al. 2014). This is also known as the Foreign-Language Effect (henceforth FLE) (Keysar, Hayakawa, and An 2012). Typically, previous work on the FLE asked participants to read moral dilemmas. Moreover, most of the FLE studies tested bilingual participants whose L1 was typologically distant from their L2. A disadvantage of relying on text and focusing on relatively distant L1–L2 language pairs is that it raises issues of generalizability of the obtained results. The aim of the current study is to test the limits of the FLE by examining (1) whether it holds for highly proficient bilinguals of a closely related language pair (i.e. Dutch-English), and (2) whether it can be replicated in an auditory setting.

Moral decision making has been defined as ‘evaluating the (good vs. bad) actions or character of a person that are made with respect to a set of virtues held to be obligatory by a culture or subculture’
(Haidt 2001, 817). Traditional theories of moral development have emphasised the role of controlled cognition in moral decision making (e.g. Kohlberg 1969), while more recent theories have highlighted the role of automatic emotional processes (e.g. Haidt 2001). Other researchers have proposed a dual-process theory (e.g. Greene et al. 2001; Kahneman 2003), according to which moral decision making is driven by a complex interaction between rational and emotional processes. In this dual-process account, controlled processes typically support judgments favouring the greater good by maximising benefits and minimising costs across affected individuals (i.e. utilitarian decisions; Mill 1861 1998), while emotional processes generally support judgments that favour the essential rights and duties of a person (i.e. deontological decisions; Kant 1785 1959).

Previous research has shown that the relative weight of rational and intuitive processes in moral decision making can vary, and lead to more or less utilitarian or deontological choices. Recent studies have shown that moral decisions depend on whether they are made on the basis of information presented in an L1 or L2 (e.g. Cipolletti, McFarlane, and Weissglass 2016; Costa et al. 2014; Geipel, Hadjichristidis, and Surian 2015b; Hayakawa et al. 2017). In Costa et al. (2014), bilinguals (i.e. Korean-English, English-Spanish, Spanish-English, English/Spanish-Hebrew, and English-French) considered the classic Footbridge dilemma (Foot 1978) in which five people tied to a train track are about to be killed by an oncoming train. The only way to save them would be to push a large bystander onto the tracks, thereby killing him but stopping the train. Only 18% of participants were willing to sacrifice the large man (utilitarian decision) when the problem was presented in their L1, whereas 44% were willing to do so when it was presented in their L2. However, this pattern was not found on the Switch dilemma. In this dilemma, a train is headed towards five workmen but participants can now choose to switch the train to another track using a lever where it would kill only one man. Pulling the lever is considered as less emotionally aversive and less personal (as it does not involve direct contact with the victim) than pushing a man on the tracks. According to Greene et al. (2008, 1146), the Footbridge dilemma is a personal dilemma because it involves (a) a serious bodily harm (b) to one or more particular individuals, where (c) this harm is not the result of deflecting an existing threat. In contrast, the Switch dilemma is an impersonal dilemma as it lacks one or more characteristics of the personal dilemma. The authors argued that thinking in an L2 creates emotional distancing compared to thinking in an L1. This reduction in emotionality might increase rational processes and thus lead to an increase in utilitarian choices. This distinctive moral decision pattern between personal and impersonal moral dilemmas has been replicated by Cipolletti et al. (2016), who studied English-Spanish bilinguals.

Although Geipel et al. (2015b) mostly replicated the results of Costa et al. (2014), they argued that the FLE was not driven by an attenuation of emotions. In Experiment 3 in Geipel et al. (2015b), they presented two additional moral dilemmas besides the personal Footbridge and the impersonal Switch dilemma to German-English bilinguals. The first dilemma was personal and high in emotion (Crying baby) and the second one was impersonal and low in emotion (Lost wallet). In the Crying baby dilemma, one must decide whether to smother one’s own child in order to save oneself and several others from being found and killed by enemy soldiers. In the Lost wallet dilemma, a person in need must decide whether to return a wallet full of cash that seems to belong to a wealthy individual.

Importantly, the FLE was found in the impersonal Lost wallet dilemma and the personal Footbridge dilemma, but it was absent in the personal Crying baby dilemma and the impersonal Switch dilemma. The authors suggested that the outcome on the Crying baby dilemma would be negative anyway (the child’s death), even if the action was not performed. The authors’ explanation why the FLE was present on the Lost wallet and the Footbridge dilemma is that these dilemmas trigger social or moral norms which prohibit people from keeping wallets or pushing other people. They therefore concluded that social or moral norms elicit the FLE and the distinction between personal versus impersonal, as suggested by Costa et al. (2014), is less important. This idea is in line with previous work revealing that disapproval of violations of everyday social and moral norms (e.g. cheating in an exam) increase in an L2 (Geipel, Hadjichristidis, and Surian 2015a; for similar views, see Dewaele...
2010; Gawinkowska, Paradowski, and Bilewicz 2013). The current study will examine more closely which type of dilemmas are driving the FLE.

Several different L1–L2 language pairs have been investigated in the prior work on the FLE. What these studies have in common is that they mainly looked at language pairs such as Spanish-English and Korean-English that are typologically distant (Cipolletti et al. 2016; Costa et al. 2014; Geipel et al. 2015b), except for Hayakawa et al. (2017) who also included one closely related language pair in their experiments (i.e. English-German and vice versa) and for Geipel et al. (2015b) who included German-English in their Experiment 3. As these studies have demonstrated an effect of language on moral decision making, it in the first instance appears to be a robust finding across a rather wide variety of language pairs. However, evidently, not all languages have been under investigation yet. The language pair tested here, which has not been studied before, will be Dutch-English. Dutch is, like English, a West-Germanic language and is one of the closest relatives of English. Dutch and English are not only linguistically but also culturally related and English has been shown to have an increasing impact in the Netherlands (Nortier 2011).

As it has been shown that amount of emotionality may influence moral decision making (e.g. Costa et al. 2014), the finding that certain kinds of words and phrases are more emotionally intense in participants’ L1 when heard than when read, is important (Caldwell-Harris and Ayçiçeği-Dinn 2009; Dewaele 2004; Harris, Ayçiçeği, and Gleason 2003; Harris, Gleason, and Ayçiçeği 2006; but see Jankowiak and Korpal 2018). For example, Harris and colleagues found that childhood reprimands, such as ‘Shame on you!’ and ‘Go to your room!’, evoked reduced skin conductance responses when they were read aloud in an L2. It may therefore be possible that Dutch-English bilinguals experience no or a reduced FLE while reading written moral dilemmas, but a strong(er) FLE while listening to spoken moral dilemmas. This idea is consistent with prior work that has demonstrated that listeners’ expectations of a speaker can influence the way language is processed and interpreted (e.g. Lev-Ari 2015; Lev-Ari and Keysar 2012). For example, listeners hold specific expectations regarding the speech of non-native speakers (Hanulikova et al. 2012). Moreover, it has been found that speakers’ foreign-accented speech can negatively impact the credibility of their message (Lev-Ari and Keysar 2010).

The present study

The main purpose of the current study is to test the robustness of the FLE. In Experiment 1, the aim was to investigate whether Dutch–English bilinguals will show a FLE when facing moral dilemmas in Dutch (L1) compared to English (L2). Costa et al. (2014) have shown that proficiency in the L2 influences the size of the FLE. More specifically, the increase in utilitarian decisions in an L2 was smaller for high than for low proficiency participants because they have developed more emotionality in their L2. Moreover, Čavar and Tytus (2018) recently showed that there was a lack of a decision-making difference in two languages in a group of highly proficient bilinguals. On the basis of these findings, it is expected that Dutch–English bilinguals, who are typically highly proficient in English, will show no or a reduced FLE. As Geipel et al. (2015b), the aim of Experiment 1 was also to shed more light on the link between the FLE and the character of moral dilemmas. To this end, the same moral dilemmas as in Experiment 3 of Geipel et al. (2015b) were used, but one additional personal and one additional impersonal dilemma was added (see Method section for more information). In this way, it was possible to examine more closely whether the personal versus impersonal distinction matters for the FLE to occur.

In Experiment 2, it will be investigated whether listening to moral dilemmas in an L2 (as opposed to in an L1) can affect decision making (henceforth auditory Foreign-Language effect). This will provide new ecologically valid insights in how moral decision takes place in everyday life situations. The results may have far-reaching consequences, as foreign languages are often used during meetings in international institutes and in international companies in general. As the difference in emotional intensity between an L1 and an L2 is accentuated during listening (Caldwell-Harris and Ayçiçeği-
Dinn 2009; Dewaele 2004; Harris et al. 2003; Harris et al. 2006), it is expected that an auditory FLE will be found, even for highly proficient bilinguals of a closely related language pair. At the same time, it is possible that listening to moral dilemmas increases participants’ cognitive load (and/or stress), because they can, as a rule (and in this experimental set-up), only be listened to once. This contrasts to reading moral dilemmas which can, in principle, be read back multiple times (unless under a certain time pressure), which gives participants ample time to consciously reflect on the dilemmas before making a moral decision. Previous work has shown that cognitive load (e.g. Cummins and Cummins 2012; Greene et al. 2008) and stress (e.g. Starcke, Ludwig, and Brand 2012; Youssef et al. 2012) influence moral decision making. More specifically, it has been found that taxing cognitive resources, decreases utilitarian decisions. It is thus possible that listening to spoken moral dilemmas, especially in the L2, increases cognitive load, which ultimately leads to a reduction in utilitarian decisions in the L2. Finally, as in Experiment 1, it will also be investigated whether the character of the moral dilemmas (personal versus impersonal) has an influence on the FLE in an auditory setting.

Experiment 1: written version

Method

Participants

Sixty Dutch-English participants (30 female, \( M_{AGE} = 27.8 \) years, \( SD_{AGE} = 12.3 \)) were recruited. Thirty participants were randomly assigned to the native language condition (Dutch) and 30 to the foreign language condition (English). Participation was voluntary and the experimental protocol was approved by the independent ethics committee of the Radboud University Nijmegen. On average, participants in the foreign language condition started learning English at the age of 10 years. They were asked to fill out their self-rated proficiency in English in terms of speaking, listening, writing and reading on a 5-point Likert scale (1 = no knowledge, 2 = beginner, 3 = average, 4 = advanced, 5 = native-like). On average, they rated their overall English proficiency skills with a score of 3.60 (SD = 0.54) and their English reading skills with a score of 3.80 (SD = 0.71).

Several independent t-tests showed that participants of the Dutch language group did not differ from those of the English language group on the self-rated linguistic background measures (listening in Dutch \( t(58) = -1.98, p = .052 \); reading in Dutch \( t(58) = -1.88, p = .07 \); listening in English \( t(58) = -2.11, p = .04 \); reading in English: \( t(58) = -1.17, p = .24 \); Age of English onset \( t(58) = -1.17, p = .24 \); English proficiency: \( t(58) = .11, p = .92 \). This reveals that, although results are based on a between-subjects design, the participants assigned to the native and foreign language groups were comparable.

Note that most of the previous work on FLE studies tested sequential bilinguals, who had acquired their L2 in a classroom setting. Although the current study included bilinguals with the same language profile, the bilingual situation of Dutch-English bilinguals is rather different from the bilinguals in the previous research, as they are known to approach fluency in their L2. More specifically, a study by the European Commission (2012) has shown that the Netherlands is one of the most multilingual nations in Europe. In a questionnaire, 94% of Dutch respondents indicated that they were able to speak at least one other language besides their mother tongue, with English the most commonly spoken of these foreign languages. Ninety percent of the respondents specified they had sufficient English in order to have a conversation. The Netherlands has thus for a reason been named the first country in the world (out of 80 countries) with the highest proficiency in the English language (Education First 2012). Nowadays, English is the official language in 74% of master studies and 23% of bachelor studies (VSNU 2018). The multilingual nature of the Netherlands, is argued to be one of the major reasons that accounts for the high level of English proficiency among Dutch inhabitants.

Materials

The same moral dilemmas were used as in Experiment 3 in Geipel et al. (2015b). They presented two personal (i.e. Footbridge, Crying baby) and two impersonal dilemmas (i.e. Switch, Lost
wallet) to participants. In the current study, one additional dilemma of each type was added (personal: Vitamins; impersonal: Taxes), resulting in six moral dilemmas in total (see Table 1 and the Appendix for complete descriptions). In the Vitamins dilemma, one must decide whether or not to remove a man’s kidney in order to save the lives of six other people. The man will not die if you do this, but his health will be compromised. In the Taxes dilemma, you are the owner of a business trying to make ends meet. One must decide whether it is appropriate to pretend that certain personal expenses are business expenses in order to lower taxes. The mean emotion ratings of these two additional dilemmas (as measured by Koenigs et al. 2007, see their supplementary material) matched closely with the mean emotion ratings of the dilemmas in Geipel et al. (2015b).

In addition to the six moral dilemmas, three filler dilemmas were presented. Note that although these dilemmas can be construed as moral, they are distinct from the experimental dilemmas because they do not directly pit a deontological versus utilitarian choice but rather a choice between self-interest and common good. The same two fillers were used (i.e. Train or bus (slightly adjusted to the Dutch situation), Plant transport) as in Geipel, Hadjichristidis, and Surian (2015b) and another self-made one was added (i.e. Spaghetti; see Appendix). The Train or bus dilemma concerned a choice between travelling by bus or train given certain time constraints. The Plant transport dilemma involved a choice between doing multiple car trips (vs. a single car trip) to avoid ruining a car’s upholstery. The Spaghetti dilemma concerned a choice between eating spaghetti leftovers versus dining out. The purpose of these dilemmas was to assess whether participants in the foreign language condition understood the English materials. The Train or bus dilemma was designed to induce a high rate of endorsements (i.e. utilitarian decisions). If there would be language misunderstandings, the endorsement rate should be around 50%. The Plant transport and Spaghetti dilemmas were concerned with self-interest (multiple car trips (polluting); going out to eat (wasting food)) versus common good (one car trip; eating leftovers). A ‘No’ answer was therefore coded as an utilitarian response (i.e. not using the car multiple times; not going out to eat).

The dilemmas were originally written in English (except for the Spaghetti dilemma). They were translated into Dutch by two native Dutch speakers who are highly proficient in English (C1 and C2 level). The translations were then compared and adjusted in consultation. The word count of the dilemmas in both languages were kept as equal as possible ($M = 103.3$ for Dutch and $M = 103.2$ for English).

### Table 1. Overview of the characteristics of the moral dilemmas.

<table>
<thead>
<tr>
<th>Type of dilemma</th>
<th>Dilemma</th>
<th>Mean emotion rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal</td>
<td>Footbridge</td>
<td>6.0 (high)</td>
</tr>
<tr>
<td></td>
<td>Crying baby</td>
<td>6.8 (high)</td>
</tr>
<tr>
<td></td>
<td>Vitamins</td>
<td>6.8 (high)</td>
</tr>
<tr>
<td>Impersonal</td>
<td>Lost wallet</td>
<td>2.9 (low)</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>5.3 (low)</td>
</tr>
<tr>
<td></td>
<td>Taxes</td>
<td>2.7 (low)</td>
</tr>
</tbody>
</table>

*As reported by Koenigs et al. (2007), on a scale from 1 to 7.

**Procedure**

The experiment was conducted in a quiet room. Participants received instructions and the moral dilemmas entirely written in one language: Dutch or English. Participants’ task was to judge the appropriateness of the proposed action by circling ‘yes’ (utilitarian decision) or ‘no’ (deontological decision) on the form. The order of dilemma presentation was randomised across participants. After this, participants filled out a (language) background questionnaire. The whole session lasted about 15 min.
Results and discussion

Data were analysed using a mixed-effects logistic regression model (Jaeger 2008) with moral decision as the binary dependent variable (0 = deontological, no; 1 = utilitarian, yes). A logistic linking function was used to deal with the categorical nature of the dependent variable. In this type of regression, the dependent variable is not directly fitted but it models the probability (in terms of logits) associated with the values of the dependent variable.

To assess the influence of language and/or the type of moral dilemmas on moral decisions, a 2 × 2 model with Language (native vs. foreign) as one contrast-coded fixed effect, Dilemma type (personal vs. impersonal) as the other contrast-coded fixed effect, and their interaction, was conducted. A second analysis focused on the filler dilemmas by conducting a model of moral decision with Language (native versus foreign) as one contrast-coded fixed effect. The maximal random effects structure included intercepts for participants but not for items, given that there were few items, which differed on purpose in many respects (see Table 1), were included in this sample. Random slopes for Language, Dilemma type and/or Norm violation by participants were added but those models never converged.

Figure 1A shows the percentage of utilitarian decisions on each dilemma. Table 2 outlines the results of the mixed-effects logistic regression model. The analysis on the moral dilemmas demonstrated no significant effect of Language, Dilemma type, or their interaction. The analysis on the filler dilemmas showed no significant effect of Language.

These results reveal that the FLE, an increase in the rate of utilitarian decisions in an L2 versus an L1, was not present. This is inconsistent with previous work which has shown a relation between moral decision making and language on personal dilemmas (Cipolletti et al. 2016; Costa et al. 2014) for several language pairs. At the same time, it is in line with the findings that highly proficient bilinguals may not or to a lesser extent experience a FLE (Čavar and Tytus 2018), as they have developed more emotionality in their L2 (Costa et al. 2014; Geipel et al. 2015b). Moreover, it is possible that the FLE does not appear for closely related language pairs, although it has been shown for German-English and English-German bilinguals (Geipel et al. 2015b; Hayakawa et al. 2017).

As previous research has shown that phrases can be more emotionally intense in participants’ L1 when heard than when read (e.g. Caldwell-Harris and Ayçiçeği-Dinn 2009; Dewaele 2004; Harris et al. 2003; Harris et al. 2006), Experiment 2 investigated whether the FLE is present in an auditory setting.

Experiment 2: auditory version

Experiment 2 replicated Experiment 1, except that the dilemmas were presented auditorily.

Method

Participants

Sixty Dutch-English participants (32 female, $M_{AGE} = 29.5$ years, $SD_{AGE} = 14.5$) were recruited. Thirty participants were randomly assigned to the native language condition (Dutch) and 30 to the foreign language condition (English). On average, participants in the foreign language condition started learning English at the age of 10 years. They were asked to fill out their self-rated proficiency in English in terms of speaking, listening, writing and reading on a 5-point scale (1 = no knowledge, 2 = beginner, 3 = average, 4 = advanced, 5 = native-like). On average, participants in the foreign language condition rated their overall English skills with a score of 3.47 ($SD = 0.49$) and their English listening skills with a score of 3.67 ($SD = 0.55$).

Several independent t-tests showed that participants of the Dutch language group did not differ from those of the English language group on most of the linguistic background measures.
There were, however, differences between the groups in their age of English onset ($t_{\text{AGEOFENGLISHONSET}}(58) = 2.71, p = .009$) and their Dutch listening skills ($t_{\text{DUTCHLISTENING}}(58) = -2.04, p = .047$). The Dutch group learnt English at a mean age of $-1.51, p = .14$; $t_{\text{ENGLISHLISTENING}}(58) = -0.01, p = 1.00$; $t_{\text{ENGLISHREADING}}(58) = .62, p = .53$; $t_{\text{ENGLISHPROFICIENCY}}(58) = .57, p = .57$.

### Table 2. Results of the glmer models for Experiment 1 (written version).

<table>
<thead>
<tr>
<th>Analysis 1: Personal versus impersonal distinction</th>
<th>Estimate</th>
<th>SE</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.62</td>
<td>0.13</td>
<td>-4.95</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Language</td>
<td>0.05</td>
<td>0.25</td>
<td>0.21</td>
<td>0.84</td>
</tr>
<tr>
<td>Dilemma type</td>
<td>0.30</td>
<td>0.22</td>
<td>1.34</td>
<td>0.18</td>
</tr>
<tr>
<td>Language:Dilemma type</td>
<td>-0.10</td>
<td>0.45</td>
<td>-0.23</td>
<td>0.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis 2: Filler dilemmas</th>
<th>Estimate</th>
<th>SE</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.38</td>
<td>0.15</td>
<td>2.52</td>
<td>0.01</td>
</tr>
<tr>
<td>Language</td>
<td>0.14</td>
<td>0.30</td>
<td>0.46</td>
<td>0.65</td>
</tr>
</tbody>
</table>
11 (SD = 1.43), whereas the English group learnt English at the age of 10.2 years (SD = 1.02). This age difference has no serious consequences for their bilingual status (i.e. between being a simultaneous or a sequential bilingual). The Dutch group rated their Dutch listening skills with a mean of 4.77 (SD = .50) and the English group with a mean of 4.97 (SD = .18). Both groups thus primarily rated themselves thus as being nativelike in Dutch.

Moreover, additional independent t-tests showed that participants in Experiment 1 did not differ from those in Experiment 2 on the linguistic background measures (age: t(118) = −.70, p = .48; listening in English: t(118) = −1.05, p = .30; reading in English t(118) = −.54, p = .59; age of English onset: t(118) = −1.22, p = .23; English proficiency: t(118) = .79; p = .43). This reveals that the experiments included comparable participant groups. Participation was voluntary and the experimental protocol was approved by the independent ethics committee of the Radboud University Nijmegen.

**Materials**

The same moral dilemmas were used as in Experiment 1. A native speaker of Dutch (male, 28 years old), who finished his BA, MA and PhD at the English Language and Culture department of the Radboud University, recorded the dilemmas (32 bit, 44,100 Hertz, using Adobe Audition©). This speaker was used for both the Dutch and the English dilemmas to prevent any influence of speaker characteristics on the results. He was chosen as our speaker because his English was of very high proficiency (C2 level). Participants in the foreign language condition were asked to fill out whether they thought that English was the native language of the speaker on a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree). On average, they rated the speaker with a score of 3.97 (SD = 1.16). None of the participants indicated that the speaker’s native language was other than English. Participants also rated the speaker to be easily understandable (M = 4.17; SD = 0.95).

**Procedure**

Participants were tested in a quiet room. They sat in front of a laptop. The experiment was presented using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Written instructions were given on the laptop screen in Dutch or English, depending on condition assignment. The moral dilemmas were presented via headphones in randomised order. Participants’ task was the same as in Experiment 1, except that they now had to press the arrow to the left on the keyboard for ’yes’ (utilitarian decision, indicated with a green sticker) or the arrow to the right for ’no’ (deontological decision, indicated with a red sticker).

Each trial had the following structure. First, a fixation cross appeared in the centre of the screen and participants heard the title of the dilemma followed by a second of silence. Then the participant heard the rest of the dilemma, followed by a second of silence before the question was asked. Immediately after this, the fixation cross disappeared and the words ’yes’ and ’no’ appeared. From that moment on, participants had 20 s to respond before the next trial initiated. Finally, participants filled out a written (language) background questionnaire. The whole session took about 15 min.

**Results and discussion**

Figure 1B presents the percentage of utilitarian decisions on each dilemma. Table 3 outlines the results of the mixed-effects logistic regression model. Similar analyses were conducted as in Experiment 1. The analysis of the data showed a significant effect of Language (estimate = .77, se = .32, z-value = 2.34, p = .02), indicating that the odds of making a utilitarian decision are significantly higher in an L1 compared to an L2 (odds ratio = 2.16). This auditory FLE effect did not interact with Dilemma type. There was also no effect of Dilemma type. The analysis on the filler dilemmas showed no significant effect of Language.
General discussion

The aim of this study was to test the limits of the FLE by examining (1) whether it holds for highly proficient bilinguals of a closely related language pair (i.e. Dutch-English), and (2) whether it can be replicated in an auditory setting.

In Experiment 1, several moral dilemmas were presented in the text to highly proficient Dutch-English bilinguals, who had to indicate whether they thought a certain action was appropriate or not. The dilemmas were either presented in Dutch or in English. It was expected that the FLE would be reduced or absent in this sample, as high proficiency in the L2 has shown to reduce (Costa et al. 2014; Geipel et al. 2015b) or eliminate the FLE (Čavar and Tytus 2018). In line with this expectation, no FLE was found, which supports the notion that increased L2 use and proficiency diminish emotional distance to the L2. In other words, high proficiency may promote emotional grounding (Costa et al. 2014). The emotional fluency in Dutch-English bilinguals may furthermore be connected to the close cultural and typological relation between Dutch and English (Nortier 2011). Until now, the FLE has mainly been found for a wide range of relatively distant language pairs, including English-Spanish, Italian-English and Korean-English (L1–L2 respectively; Cipolletti et al. 2016; Costa et al. 2014; Geipel et al. 2015b). As the current results show that in some cases (i.e. high proficiency, closely related language pairs) individuals make more/less use of intuitive emotionally driven processes and/or rational conscious processes, they fit well with the dual-process account of moral decision making (e.g. Greene et al. 2001; Greene and Haidt 2002; Kahneman 2003).

In Experiment 2, highly proficient Dutch-English bilinguals were presented with spoken moral dilemmas. Their task was similar to Experiment 1. It was predicted that the FLE would be present, as previous research has shown that linguistic information can be more emotional in the L1 when heard than when read (e.g. Caldwell-Harris and Ayçiçeği-Dinn 2009; Dewaele 2004; Harris et al. 2003; Harris et al. 2006). The results confirmed this prediction: an auditory FLE was found. Individuals make more utilitarian decisions when hearing a moral dilemma in their L2 than in their L1. The effect of language did not interact with dilemma type (personal versus impersonal). This finding goes against the hypothesis that the FLE is constrained to personal dilemmas and the amount of emotionality involved (see Costa et al. 2014). Note, however, that Costa et al. (2014) only collected data on the Footbridge and the Switch dilemma using a reading task. Comparing those data with the current reading data on the Footbridge and the Switch dilemma (Experiment 1) shows that the percentage of utilitarian decisions in the L1 and the L2 are rather similar across the two studies.

If emotional attenuation is not the conclusive explanation for the (auditory) FLE on moral decision making, the question remains what drives this effect? Some researchers have suggested that an L2 may influence moral decision making by reducing the mental accessibility of social and moral rules (Geipel, Hadjichristidis, and Surian 2015a; Geipel et al. 2015b; Dewaele 2010; Gawinkowska, Paradowski, and Bilewicz 2013). For example, Geipel and colleagues found that the FLE was present in the Footbridge (personal) and the Lost wallet dilemma (impersonal), but not in the Switch dilemma (impersonal). The authors argued that the first two dilemmas involve a prohibited action (pushing people, keeping lost wallets), whereas the third dilemma does not
(there is no general rule which prohibits people from flipping switches). The current data seem to be in line with this reasoning when only looking at the percentage of utilitarian decisions in the L1 versus the L2 on these dilemmas (see Figure 1). However, future work needs to more systematically study this by including an equal amount of norm versus no-norm violation dilemmas. Such a study would be able to confirm whether this distinction is truly driving the (auditory) FLE.

An anonymous reviewer pointed out that for almost all dilemmas, in particular in Experiment 2, the percentage of utilitarian responses in the L2 versus L1 was closer to 50% (with the exception of the Switch dilemma). This could indicate that when people use an L2 they become more uncertain, and are therefore more likely to go for a 50–50 approach (see also Geipel et al. 2015b; Muda et al. 2018). Alternatively, it could mean that the current participants found the dilemmas harder to understand in the L2 than in the L1 and they were thus not as highly proficient in their L2 as assumed. However, this 50–50 L2 approach did not hold for the filler items in both experiments. In both cases, the response mean in the L1 was closer to 50% than the response mean in the L2.

The current research extends the previous work by revealing that listening to moral dilemmas could elicit a FLE in highly proficient bilinguals. These results fit with previous work which has shown that foreign-accented speech can negatively impact the reliability of a message (Lev-Ari and Keysar 2010). Here, it has been demonstrated that hearing a foreign language can even influence moral decision making, and namely promote more utilitarian-type decisions. This may have societal consequences, as important international decisions, such as those taken at the United Nations, often take place in auditory settings.

Note, however, that it was not possible to directly compare the written and auditory experiment, as there were also other differences between the two set-ups (e.g. timing of presentation, ability to review the presentation). This makes a direct comparison between these two experiments less valid. Future research should take this into account by manipulating mode of presentation in a single experiment. Such an experiment would, for example, also be able to assess whether the auditory condition elicits more utilitarian decisions in general than the visual condition. As the results of each experiment independently are so different from each other, it is suggested though that the current findings may provide evidence for the situational ethics approach (Fletcher 1997). This view holds that the morality of an act is determined by its context rather than by judging it according to absolute moral standards. In this case, the auditory context made individuals respond in a more utilitarian fashion in their L2, whereas this did not occur in a written context.

There are a couple of limitations to this study. Although more moral dilemmas (six in total) were used than in most of the previous work on the FLE, it is still a rather restricted number. Future research should therefore present a wider variety of dilemmas. Furthermore, some of the dilemmas that have been presented are rather distant from real life (e.g. Footbridge), although the auditory FLE was also found on a more ordinary example (e.g. Lost wallet). It is important that future research incorporates more realistic dilemmas like in Geipel et al. (2015a). Finally, although the number of participants matches more or less with the number tested in Geipel et al. (2015b), other studies on moral decision making have often included more participants but those studies presented fewer moral dilemmas.

In light of the current study, a number of ideas for future research could be proposed. First, it would be interesting to directly compare typologically close (e.g. Dutch-English) to typologically distant language pairs (e.g. Korean-English), while controlling for factors such as L2 proficiency. Secondly, it would be relevant to systematically test the role of proficiency on the strength of the (auditory) FLE by including groups of L2 learners who differ in their proficiency levels. It is thereby important to compare L2 learners with the same L1 and L2 to control for linguistic or cultural differences. In the current study, most bilinguals self-rated themselves as advanced in English (around 70%). It would be interesting for follow-up research to compare a group of participants with more varying degrees of proficiency to draw stronger conclusions on the effect of proficiency. Besides collecting self-rated linguistic measures, as in the current study, it would be informative to also use norm-referenced measures to assess participants’ L2 proficiency level. Finally, it would be important
to further test the auditory FLE with several other language pairs to see if the current results can be replicated.

In conclusion, the current results show that the FLE is absent in the current sample of highly proficient Dutch-English bilinguals read moral dilemmas, but it is present when they listen to moral dilemmas. These findings aid in understanding the robustness of the FLE, revealing that in some contexts it may be overcome and/or inhibited, whereas in others it may be strengthened.

Note

1. The Dutch-English language pair has been tested in previous studies but not in the context of moral judgment. For example, Puntoni, De Langhe, and Van Osselaer (2009) used Dutch-English bilinguals to study the emotionality of advertising messages, Geipel, Hadjichristidis, and Klesse (2018) studied willingness to consume sustainable but disgusting products, and Urbig et al. (2016) studied free riding in a group context.

Acknowledgements

I would like to thank Marieke Korsten, Marije van’t Kruijs, Julia Merkus, and Esther Vullings for their enthusiasm, their assistance with participant running, and the fruitful discussions. I would also like to thank Ruud van den Beuken for lending his voice.

Disclosure statement

No potential conflict of interest was reported by the author.

Data availability statement

There is no DOI available. The author could add the data sets as supplementary materials.

References


Appendix: Dilemmas used in Experiment 1 and 2

(English versions, mean emotion rating (on a scale from 1 to 7) is reported by Koenigs et al. 2007).

**Personal dilemmas**

Footbridge (mean emotion rating: 6.0). A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved. Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?

Crying baby (mean emotion rating: 6.8). Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others you must smother your child to death. Is it appropriate for you to smother your child in order to save yourself and the other townspeople?

Vitamins (mean emotion rating: 6.8). You are the leader of a mountaineering expedition that is stranded in the wilderness. Your expedition includes a family of six that has a genetically caused vitamin deficiency. A few people’s kidneys contain large amounts of this vitamin. There is one such person in your party. The only way to save the lives of the six members of this family is to remove one of this man’s kidneys so that the necessary vitamins may be extracted from it. The man will not die if you do this, but his health will be compromised. The man is opposed to this plan, but you have the power to do as you see fit. Is it appropriate for you to forcibly remove this man’s kidney in order to save the lives of the six vitamin-deficient people?

**Impersonal dilemmas**

Switch (or Trolley) (mean emotion rating: 5.3). You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. Is it appropriate for you to hit the switch in order to save the lives of the five workmen?

Lost wallet (mean emotion rating: 2.9). You are walking down the street when you come across a wallet lying on the ground. You open the wallet and find that it contains several hundred euros in cash as well the owner’s driver’s license. From the credit cards and other items in the wallet it’s very clear that the wallet’s owner is wealthy. You, on the other hand, have been hit by hard times recently and could really use some extra money. You consider sending the wallet back to the owner without the cash, keeping the cash for yourself. Is it appropriate for you to keep the money you found in the wallet in order to have more money for yourself?

Taxes (mean emotion rating: 2.7). You are the owner of a small business trying to make ends meet. It occurs to you that you could lower your taxes by pretending that some of your personal expenses are business expenses. For example, you could pretend that the stereo in your bedroom is being used in the lounge at the office, or that your dinners out with your partner are dinners with clients. Is it appropriate for you to pretend that certain personal expenses are business expenses in order to lower your taxes?

**Filler dilemmas**

Plant transport. You are bringing home a number of plants from a store that is about 5 km from your home. The trunk of your car, which you’ve lined with plastic to catch the mud from the plants, will hold most of the plants you’ve purchased. Is it appropriate for you to make two trips home in order to avoid ruining the upholstery of your car?

Train or bus. You need to travel from Nijmegen to Amsterdam in order to attend a meeting that starts at 2:00 pm. You can take either the train or the bus. The train will get you there just in time for your meeting no matter what. The bus is scheduled to arrive an hour before your meeting, but the bus is occasionally several hours late because of traffic. It would be nice to have an extra hour before the meeting, but you cannot afford to be late. Is it appropriate for you to take the train instead of the bus in order to ensure you are not being late for your meeting?
Spaghetti. You have decided that you want to live less wasteful. Yesterday you made spaghetti with fresh ingredients for yourself for two days. Now that you are home you do not want to eat the spaghetti. If you do not eat it today, you can eat it tomorrow. Is it appropriate for you to throw away the spaghetti and pick up something you do feel like eating instead of the spaghetti?