

## The affect of negativity: testing the Foreign Language Effect in three types of valence framing and a moral dilemma

Bregje Holleman, Naomi Kamoen & Marijn Struiksma

To cite this article: Bregje Holleman, Naomi Kamoen & Marijn Struiksma (2021) The affect of negativity: testing the Foreign Language Effect in three types of valence framing and a moral dilemma, *Cognition and Emotion*, 35:4, 690-704, DOI: [10.1080/02699931.2021.1889472](https://doi.org/10.1080/02699931.2021.1889472)

To link to this article: <https://doi.org/10.1080/02699931.2021.1889472>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 23 Feb 2021.



[Submit your article to this journal](#)



Article views: 894






[View related articles](#)



[View Crossmark data](#)

# The affect of negativity: testing the Foreign Language Effect in three types of valence framing and a moral dilemma

Bregje Holleman <sup>a</sup>, Naomi Kamoen <sup>b</sup> and Marijn Struiksmā <sup>a\*</sup>

<sup>a</sup>Utrecht Institute of Linguistics Uil OTS, Utrecht University, Utrecht, Netherlands; <sup>b</sup>Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

## ABSTRACT

In decision-making people react differently to positive wordings than to negatives, which may be caused by negativity bias: a difference in emotional force of these wordings. Because emotions are assumed to be activated more strongly in one's mother tongue, we predict a Foreign Language Effect, being that such framing effects are larger in a native language than in a foreign one. In two experimental studies ( $N=475$  and  $N=503$ ) we tested this prediction for balanced and unbalanced second language users of Spanish and English and for three types of valence framing effects. In Study 1 we observed risky-choice framing effects and attribute framing effects, but these were always equally large for native and foreign-language speakers. In our second study, we added a footbridge dilemma to the framing materials. Only for this task we did observe a Foreign Language Effect, indicating more utilitarian choices when the dilemma is presented in L2. Hence, across two studies, we find no Foreign Language Effect for three types of valence framing but we do find evidence for such an effect in a moral decision task. We discuss several alternative explanations for these results.

## ARTICLE HISTORY

Received 16 July 2021  
Revised 27 January 2021  
Accepted 8 February 2021

## KEYWORDS

Valence framing; negativity bias; Foreign Language Effect (FLE); footbridge dilemma; risky-choice framing; L1/L2


## Introduction

When people make choices, they should not be influenced by irrelevant factors such as variations in the wording describing the available options. Yet, research has repeatedly demonstrated that people systematically express different preferences based on the same scenarios depending on how the options or outcomes are described linguistically (Kahneman, 2003). Likewise, communication researchers and methodologists have found that people's responses to survey questions can vary due to subtle differences in question form and wording (Schuman & Presser, 1981). Understanding the causes of framing effects and the contexts in which they occur is relevant to everyday communication,

and can inform theories of language processing and decision-making.

Valence framing effects, or differences in how people respond to equivalent information depending on its positive or negative wording, have often been ascribed to emotions (De Martino et al., 2006). In risky-choice framing tasks, for example, people make riskier decisions in a negative frame in order to maximise their chances on loss avoidance, whereas they prefer a more certain decision when the focus is on a possible positive outcome in order to secure their gain (Kühberger, 1998; Tversky & Kahneman, 1986). As to the cause of this asymmetry, the risk-as-feelings hypothesis (Loewenstein et al., 2001) suggests that in decision-making under risk and uncertainty, not only

**CONTACT** Bregje Holleman  [b.holleman@uu.nl](mailto:b.holleman@uu.nl),  Department of Languages, Literature and Communication, Utrecht Institute of Linguistics Uil OTS, Utrecht University, Trans 10, Utrecht 3512 JK, Netherlands  [@bregjeholleman](https://twitter.com/bregjeholleman)  
\*We thank Emily Felker, MA, who developed and administered the first study, which was presented in an earlier version during the conference of the International Communication Association (2018).

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/02699931.2021.1889472>

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

cognitive evaluations, but also feeling states come into play. While cognitive evaluations of risk are influenced by probabilities and outcome valence, emotional reactions to risk are influenced by a range of other factors, such as the vividness of the mental imagery evoked. According to Loewenstein et al. (2001), when emotional reactions and cognitive appraisals diverge from one another, the emotional reactions often win out when it comes to determining behaviour.

The difference in (emotional) reactions to positive and negative words is not specific to risky-choice framing, but also occurs in other valence framing situations in various contexts such as question answering or evaluation. The cause of these framing asymmetries would be in the difference in extremity and emotionality of the positive vs. the negative wording. This is described by negativity bias (Baumeister et al., 2001; Rozin & Royzman, 2001), the notion that “bad is stronger than good” (Baumeister et al., 2001) in the way people perceive and respond to events.

In language, negativity bias describes negative words to be generally more marked and to show a more extreme deviation from neutrality in their semantic meanings than their positive counterparts (Rozin et al., 2010). Negativity bias suggests automatic emotion activation is stronger in response to negative words than to positive words. Hence, reasoning from negativity bias, asymmetries in responses to various types of negative versus positive language materials, such as attribute framing or question polarity effects, would be comparable to the causes of the asymmetry in reactions to risky-choice frames.

The emotion-based explanation for framing effects was the foundation for recent studies that compared risky-choice framing effects and other decision biases in a native language (L1) and a foreign language (L2). These studies found that risky-choice framing effects were substantially reduced in an L2 (Costa et al., 2014a; Keysar et al., 2012), which was ascribed to weaker emotional reactivity in an L2 relative to an L1. We will discuss this body of literature below (Section 1.1). These studies investigated several decision tasks and only one type of valence framing: risky-choice framing. Although risky-choice framing tasks are quite common research paradigms, they are unusual decision-making tasks in daily-life situations. Therefore, in the present study, we extend this line of research by examining Foreign Language Effects across multiple framing contexts. In Section 1.2, we will discuss literature on the two additional types of framing effects

included in our studies and emotions as a possible underlying mechanism.

### **1.1 Decision making and emotion activation in a foreign language**

If framing effects and other decision biases result from affective heuristics running counter to analytical reasoning processes (Kahneman, 2003), these effects should be reduced when using a second language – as automatic emotion activation is weaker in an L2 than in L1 (Pavlenko, 2012). Indeed, in L2, people show less loss aversion when being offered a series of bets with positive expected value (Keysar et al., 2012), they show a reduction in psychological accounting biases (Costa et al., 2014a), and they are no longer susceptible to the “hot hand” fallacy when gambling (Gao et al., 2015).

Earlier studies suggest different underlying mechanisms as to why language users would show weaker emotional responses in their L2 (Dewaele, 2004; Eilola & Havelka, 2010; Harris et al., 2003). One option is that processing and responding in one’s L2 is less automatic and takes a more deliberate state of processing, resulting in a weaker emotional context when using an L2, i.e. an increased-systematicity account for L2 (Degner et al., 2012; Opitz & Degner, 2012). This may be due to cognitive load (Hadjichristidis et al. 2017): L1 users showed similar patterns to L2 users once the perceptual load in L1 was increased (Thoma & Baum, 2019).

Another explanation is that emotions are activated more strongly in one’s L1 because it is acquired at a younger age, synchronously to acquiring morality. This would lead to a much stronger embodiment of emotions in one’s L1 than in L2, i.e. an enhanced-emotionality account for L1 (Harris et al., 2003; Iacozza et al., 2017). Converging evidence from behavioural, psychophysiological, and neuroimaging studies support the idea of “disembodied cognition” in L2 (Pavlenko, 2012). That is, when an L2 is processed semantically, it may not be processed emotionally to the same degree as an L1. As a result, emotion-laden words in L2 do not always rapidly capture attention during automatic word processing, i.e. they do not show the typical negativity bias (Colbeck & Bowers, 2012), and they often lead to less physiological arousal as measured by skin conductance (Eilola & Havelka, 2010; Harris et al., 2003, 2006). In addition, ERP evidence by Jończyk et al. (2016) suggested that negatively valenced L2 words do not show full

semantic integration in early processing stages. This processing disadvantage for L2 words with negative valence might contribute to L1-L2 differences in framing effects, given that framing effects depend on a stronger emotional reaction to the negative frames.

Keysar et al. (2012) were the first to demonstrate that risky-choice framing effects are reduced in L2. In their study, participants read variants of the Asian disease problem in their L1 or in their L2. While participants showed the expected risk-aversion in the gain frame and risk-seeking in the loss frame in their L1, this asymmetry in risk preference between frames disappeared among those who completed the task in their L2. This finding was replicated by Costa et al. (2014a, 2015) and by Winkler et al. (2016). These studies provide converging evidence that people seem to be less susceptible to decision-making biases linked to automatic emotion activation when completing tasks in a foreign language.

Reasoning from this body of literature, we expect to find a Foreign Language Effect (FLE) in decision-making, i.e. weaker decision biases in L2 than in L1. We focus predominantly on a replication of FLE in risky-choice framing, and broaden that to two other types of valence framing: attribute framing and question polarity in political attitude surveys.

## 1.2 Valence framing and emotions

Emotions activated by positive and negative words and the difference in extremity and emotion of negative words (negativity bias) may well be an underlying cause of risky-choice framing effects and other framing effects of positivity and negativity. This was already elaborated on for risky-choice framing, but not for the two other valence framing effects we focus on in our studies.

*Attribute framing.* The description of a single attribute of an object or event in either positive or negative terms can affect how people evaluate the attractiveness of the item as a whole (Levin et al., 1998). To measure these attribute framing effects, half a group of participants view a description of an object, person or event in a positive frame (e.g. “This course has a 90% success rate”), the other half view it in the negative frame (“[...] a 10% failure rate”), and the mean evaluation of the attitude object is compared between the two framing conditions. Attribute framing effects have been shown to occur in a variety of contexts, from consumer evaluations of products to medical decision-making (e.g.

Holleman & Maat, 2009; Levin & Gaeth, 1988). Studies show a robust “valence-consistent shift”: items described in positive terms are rated more favourably than items described in negative terms.

To explain this phenomenon, Levin and Gaeth (1988) were the first to take an information processing approach, arguing that reading about positive or negative attributes of an item tends to evoke positive or negative associations. Levin et al. (1998) have equated this effect to valence-based priming: encoding a positive or negative attribute makes it easier to access either positively – or negatively – valenced knowledge. This account is compatible with research on affective priming, showing that words with strong emotional valence often activate evaluative reactions in a rapid, automatic way (Fazio, 2001; Klauer & Musch, 2003) and that the automatic activation of an emotional response can influence later decisions (Winklerman et al., 2007). Hence, attribute framing effects may also be related to emotional reactions to positive and negative stimuli and potentially susceptible to an FLE.

*Question polarity in political attitude surveys.* Valence framing effects also occur when people express their own attitudes about issues in response to survey questions. A well-known example of a framing effect in the context of attitude surveys is the so-called “forbid/allow asymmetry”. In an oft-cited experiment by Rugg (1941), people were asked either “Do you think the United States should forbid public speeches against democracy?” or “Do you think the United States should allow public speeches against democracy?” This experiment and many replications show that respondents are more willing to say “no” to negative wordings (i.e. forbidding) than to say “yes” to “positive ones” (i.e. allowing; also see Holleman et al., 2016; Kamoen et al., 2013; Schuman & Presser, 1981).

As with other types of valence framing effects, emotional responses evoked by the critical words, especially the negative ones, may play a role. For instance, Schuman and Presser (1981; see also Holleman, 2000) speculate that the “tone of wording”, such as the harsh sound and connotations of the word “forbid”, might contribute to the forbid/allow asymmetry. That is, endorsing the negative word feels as though one takes a more extreme position than it is to answer “no” to its positive counterpart.

The forbid/allow asymmetry could also potentially be explained by the negativity bias. Negatively valenced words have been shown to capture attention rapidly and for a longer time than positive words, reflecting automatic vigilance for negative

stimuli (Estes & Adelman, 2008). Given the evolutionary origin of the negativity bias as a way to protect against threats (Taylor, 1991), the reluctance to endorse negatively-worded questions might reflect a subconscious attempt to distance oneself from negative stimuli. Crucially, the drive to avoid negative stimuli is often stronger than the drive to approach positive stimuli. So, negative words like “forbid” might quickly trigger negative appraisals and negative action tendencies that predispose people to disagree. Because of the negativity bias, the effect of negative emotional activation may be more powerful than the corresponding positive emotion route, in which positive words would make people more inclined to agree. Therefore, we expect to find an FLE for question polarity effects too.

### 1.3 The present study

We test Foreign Language Effects for three types of valence framing, hypothesising smaller framing effects in L2 compared to L1. We predict a risky-choice framing effect in a decision making task (showing people to be more risk-averse in the positive frame and more risk-seeking in the negative one), a valence-consistent shift in attribute framing tasks in informative texts (positively-worded descriptions will lead to higher favourability ratings than logically equivalent negative statements). In response to opinion questions worded positively or negatively, we predict that negatively worded statements will obtain more negative answers compared to their positive counterparts.

For risky-choice framing, an FLE has repeatedly been established (e.g. Costa et al., 2014a; Keysar et al., 2012). However, FLE research is usually conducted with one language as the L1 and another language as the L2. In a moral decision task, Costa et al. (2014b) used a symmetrical design, in which the materials were offered in Spanish and English, and the group of participants was such that either Spanish or English would be people’s L1 or L2. We expect to replicate the FLE for risky-choice framing using a symmetrical design, similar to Costa et al. (2014b), allowing an investigation of the FLE twice in the same study.

## 2. Method Study 1

### 2.1 Design

An experimental study with a 3 (Language user: Native English with Spanish as L2, Native Spanish

with English as L2, Balanced English-Spanish Bilinguals) x 2 (Survey Version: 1 or 2, both with a mix of positively and negatively worded framing dilemmas of the three types) x 2 (Language of the study: Spanish or English) between-subjects design was implemented. The research institute’s ethical research committee approved this design (Holle102-02-2018).

### 2.2 Participants

To be eligible to take part in our online survey, participants had to speak both Spanish and English, and at least one of these languages had to be acquired from birth. To reach these groups, we recruited participants via Facebook groups and Reddit forums for Spanish language learners, as well as via MTurk: a crowdsourcing marketplace where people complete short tasks in exchange for monetary compensation.

In total, our survey was completed by 489 participants. The data of 14 participants had to be discarded, because they did not learn either Spanish or English from birth onwards. Of the eligible respondents, we classified two groups of unbalanced bilinguals: 207 participants were native speakers of English with Spanish as L2 (Mean age = 30.86;  $SD = 11.41$ ; Mean age of acquiring Spanish = 14.38,  $SD = 7.40$ ), and 146 participants were native speakers of Spanish with English as L2 (Mean age = 35.52;  $SD = 10.04$ ; Mean age of acquiring English = 8.43,  $SD = 5.05$ ). A native speaker of a language (L1) was defined to be someone who had acquired that language from birth, whereas an L2 was acquired any time after that. Additionally, we distinguished a group of 122 balanced bilinguals who had learned both Spanish and English from birth (Mean age = 29.80;  $SD = 9.08$ ). The three groups of language learners were comparable with respect to gender ( $\chi^2(2) = 5.57$ ;  $p = .06$ ), educational level ( $\chi^2(4) = 5.29$ ;  $p = .26$ ), the number of people assigned to the two versions of the experimental survey ( $\chi^2(2) = 2.89$ ,  $p = .24$ ), and the number of participants assigned to the English and the Spanish version ( $\chi^2(2) = 0.55$ ;  $p = .77$ ).

We also checked whether our three groups of language learners have different languages to automatically express strong emotions in, and feel that a certain language has a greater emotional impact on them. For both questions, we find clear differences between the three groups ( $\chi^2(8) = 256$ ;  $p < .001$  and  $\chi^2(8) = 121.29$ ;  $p < .001$ ) indicating that native speakers of Spanish prefer Spanish for the expression of

emotions whereas native speakers of English have a preference for English.

With these definitions of L1 and L2, the 475 participants were distributed over the cells of our  $3 \times 2 \times 2$  design with at least 27 participants in each cell, which is comparable to the risky-choice dilemma study in Keysar et al. (2012). A power analysis revealed that a medium or large effect size can be assessed with 99% certainty based on 475 participants.<sup>1</sup>

### 2.3 Materials and procedure

We constructed a survey in two different versions, each version first containing question polarity items, followed by attribute framing items and concluding with a risky-choice framing item about political issues. This way, a coherent survey was created that did not reveal that the actual goal of the study was to test framing effects. Participants were randomly assigned to one of the survey version in English or Spanish; for details on the procedure to test the language proficiency of participants prior to entering the study, see Appendix 1 (online).

The survey started with question polarity items ( $N = 20$ ) that were taken from two popular Voting Advice Applications, Election Compass (2016), and ISideWith (2016). There already was an English and Spanish version of the statements in Election Compass. These original wordings were used as much as possible. The twenty question items offer statements with viewpoints relevant to the 2016 American presidential elections in order to help citizens compare their own opinions with the political parties in the election. Each statement expressed a viewpoint about some political issue relevant to the 2016 American presidential elections (e.g. abortion, taxes, and foreign policy issues). For each statement, there were two possible versions that each expressed the exact opposite position (see Appendix 2 (online)). Participants had to respond to each statement by selecting “Strongly Disagree”, “Disagree”, “Neutral”, “Agree”, or “Strongly Agree”. The positive and negative versions of each statement were distributed across the two versions of the survey in such a way that each participant saw each statement only once (in either positive or negative form), and that each participant saw an equal amount of positive and negative statements (see Appendix 2 (online)).

Next, participants were confronted with eight candidate evaluation questions to examine attribute framing effects. These tasks were developed in

English, and then translated into Spanish by a Latin-American native speaker of Spanish with expertise in linguistics and language teaching. This translation was proofread by a second native Spanish speaker. Each attribute framing dilemma consisted of a brief description of a hypothetical political candidate running for a certain office along with some aspect of his or her prior experience. The last sentence of each description was a fact about the candidate’s record expressed in numerical terms. The same fact could be framed in either a positive or a negative way (see Appendix 3 (online)). Based on the little information available in the description, participants had to give their evaluation of each candidate by choosing “Very Unfavorable”, “Unfavorable”, “Neutral”, “Favorable”, or “Very Favorable”.

Finally, a variant of the Asian disease problem created by Kahneman and Tversky (1979) and adapted by Keysar et al. (2012) was presented. Participants were instructed to put themselves in the position of a policymaker and to decide on the best course of action to take, see Appendix 4 (online). In both cases, the expected value of choosing Programme A or Programme B is equal, so the choice represents participants’ preference for a sure option versus a risky gamble.

After the three framing tasks, demographic questions were posed. Most questions were based on the Language Experience and Proficiency Questionnaire (Marian et al., 2007), a validated tool for assessing the language profile of multilingual adult populations. To determine whether the different groups of L2 speakers had the expected different levels of emotional grounding in Spanish and English, the questionnaire also asked about participants’ preferred language for expressing emotional content and which language they perceived as more emotional (inspired by the Bilingual Emotionality Questionnaire; Dewaele & Pavlenko, 2001–2003). At the end of the survey, participants rated their overall comprehension of the survey materials and they could optionally provide feedback about the survey.

### 2.4 Data analyses

To test our hypotheses, the data for each type of framing were analysed separately. For the eight attribute framing tasks, as well as for the twenty survey questions, we performed two separate multilevel model analyses using MLWin (v2.34). The fixed parts of these models were built additively, which means



that we started out from an empty model with only one constant, and subsequently tested whether variation in frame, language, the participant group, all possible two-way interactions, and the three-way interaction between these variables improved the fit of the model. The result of this exercise is one model that fits the data best. As for the random part of this model, we allowed the scores to vary between participants, between items, and due to the interaction between participant and item. The item- and participant variance were estimated simultaneously, resulting in a cross-classified model (Quené & van den Bergh, 2004, 2008, see Appendix 5 (online) for details on the final models). In such a cross-classified multi-level model, effect sizes can be expressed in different ways. We decided to report an austere measure for the effect sizes qualifying all statistically significant effects relative to the squared sum of the three variance components in our models. Hence, our measure for the effect size may be considered a strict version of Cohen's  $d$  (Cohen, 1988). In terms of labelling, a Cohen's  $d$  between 0.2 and 0.5 can be considered a small effect, a Cohen's  $d$  between 0.5 and 0.8 is a medium-sized effect and a Cohen's  $d$  larger than 0.8 can be called large.

As there was only one item measuring risky-choice framing, this type of framing was analysed in a unilevel model allowing only between-participant variance. Moreover, due to the binomial nature of the dependent variable in this task, we used a Logit model. Similar to attribute framing and question framing, we built the model for risky-choice framing additively. To facilitate the interpretation of the effects for these binomial models, we report the Standardised Effect score as a measure of effect size (Van den Bergh, 1990), which is estimated by dividing the effect by the large-sample variance, which is the pooled variance for all the cells in a contingency table (see Appendix 6 (online) for an explanation). The resulting Standardised Effect ( $SE$ ) is a  $z$ -score with an associated  $p$ -value. The  $SE$  is interpreted in relation to the random sample fluctuations, i.e. a significant  $SE$  of 4 indicates that the effect is 4 times larger than the sample fluctuations.

### 3. Results study 1

Only the results of the final model are discussed, as this is the model that fits the data best. Although this final model includes different parameters for each of the three framing types, for reasons of

presentational clarity and comparability, we include a similar table with mean scores for all experimental conditions, for each type of framing. Appendix 5 (online) contains additional information about the way the final models were fitted.

*Risky-choice framing.* Table 1 shows the percentage of choices for the sure option (200 lives saved and 400 lives lost) in all experimental conditions. Results demonstrate that the choice for the sure option depends on the positive or negative frame ( $z = 9.72$ ;  $p < .001$ ;  $SE = 8.67$ ;  $p < .001$ ). In line with previous research, we found that the sure option is chosen more often in the gain-frame condition than in the loss-frame condition. The  $SE$  indicates that the framing effect is almost nine times larger than the random sample differences. Besides this main effect of framing, the only additional factor that significantly improved the model was a three-way interaction between the frame, the language of the survey, and the language user type ( $z = 2.53$ ;  $p < .01$ ;  $SE_{L1/L2 \text{ English, L1/L2 Spanish}} = -0.08$ ;  $p = .94$ ;  $SE_{Bilinguals} = -2.88$ ;  $p = .004$ ). This interaction shows that the preference for the sure option for gain-frames as compared to loss-frames was somewhat smaller for the balanced bilinguals who were assigned to the Spanish survey version; for this specific group, the difference was "only" about 25%, whereas for all other groups the difference reached up to almost 50%. To interpret the result of the three-way interaction we computed two  $SE$  scores (see Appendix 6 (online) for details). The first indicates that the effect of framing in the English and Spanish Survey was similar in the L1/L2 English and L1/L2 Spanish groups. The second  $SE$  reflects the framing effect in Bilinguals compared between the languages of the survey and indicates that for the Bilinguals the framing effect in the Spanish survey is almost 3 times smaller than that in the English survey.

*Attribute framing.* Results demonstrate a main framing effect across the board in all languages and language learner groups ( $z = 5.70$ ;  $p < .001$ ; Cohen's  $d = 0.18$ ; see Table 2 for the means): candidates described with positive frames receive more positive evaluations. Moreover, results indicate a main effect of the language in which the survey was presented: on average, the candidates receive higher evaluations when they are described in Spanish rather than in English ( $z = 3.96$ ;  $p < .001$ ; Cohen's  $d = 0.18$ ).

The framing effect is larger when a participant read the survey in Spanish rather than in English ( $z = 4.26$ ;  $p < .001$ ; Cohen's  $d = 0.24$ ), and also when the

**Table 1.** Proportion of choices for the sure-option (*SD* between brackets) for English L1, Spanish L1 and Balanced Bilinguals (in short: bilinguals).

|                | English survey ( <i>N</i> =248) |                               |                               |                | Spanish survey ( <i>N</i> =227) |                               |                               |                | Total                          |                                |                                |                |
|----------------|---------------------------------|-------------------------------|-------------------------------|----------------|---------------------------------|-------------------------------|-------------------------------|----------------|--------------------------------|--------------------------------|--------------------------------|----------------|
|                | English L1<br>( <i>N</i> =103)  | Spanish L1<br>( <i>N</i> =69) | Bilinguals<br>( <i>N</i> =63) | Total          | English L1<br>( <i>N</i> =104)  | Spanish L1<br>( <i>N</i> =77) | Bilinguals<br>( <i>N</i> =59) | Total          | English L1<br>( <i>N</i> =207) | Spanish L1<br>( <i>N</i> =146) | Bilinguals<br>( <i>N</i> =122) | Total          |
| $M_{Positive}$ | 0.82 (0.39)                     | 0.87 (0.34)                   | 0.82 (0.40)                   | 0.83<br>(0.38) | 0.69 (0.47)                     | 0.77 (0.43)                   | 0.76 (0.44)                   | 0.73<br>(0.45) | 0.75 (0.44)                    | 0.82 (0.39)                    | 0.79 (0.41)                    | 0.78<br>(0.42) |
| $M_{Negative}$ | 0.39 (0.49)                     | 0.29 (0.46)                   | 0.17<br>(0.38)                | 0.29<br>(0.46) | 0.26<br>(0.44)                  | 0.34<br>(0.48)                | 0.53 (0.51)                   | 0.36<br>(0.48) | 0.33 (0.47)                    | 0.32<br>(0.47)                 | 0.33 (0.48)                    | 0.33<br>(0.47) |
| $M_{Total}$    | 0.61<br>(.49)                   | 0.61<br>(0.49)                | 0.44<br>(0.50)                | 0.57<br>(0.50) | 0.51<br>(0.50)                  | 0.56<br>(0.50)                | 0.64<br>(0.48)                | 0.56<br>(0.50) | 0.56 (0.50)                    | 0.58 (0.50)                    | 0.54 (0.50)                    | 0.56<br>(0.50) |

**Table 2.** Mean candidate evaluations (*SD* between brackets) for English L1, Spanish L1 and Balanced Bilinguals (in short: bilinguals) (*SD* between brackets).

|                | English survey ( <i>N</i> =248) |                               |                               |                | Spanish survey ( <i>N</i> =227) |                               |                               |                | Total                          |                                |                                |                |
|----------------|---------------------------------|-------------------------------|-------------------------------|----------------|---------------------------------|-------------------------------|-------------------------------|----------------|--------------------------------|--------------------------------|--------------------------------|----------------|
|                | English L1<br>( <i>N</i> =103)  | Spanish L1<br>( <i>N</i> =69) | Bilinguals<br>( <i>N</i> =63) | Total          | English L1<br>( <i>N</i> =104)  | Spanish L1<br>( <i>N</i> =77) | Bilinguals<br>( <i>N</i> =59) | Total          | English L1<br>( <i>N</i> =207) | Spanish L1<br>( <i>N</i> =146) | Bilinguals<br>( <i>N</i> =122) | Total          |
| $M_{Positive}$ | 3.30 (1.02)                     | 3.47 (0.95)                   | 3.48 (0.98)                   | 3.40<br>(1.00) | 3.40 (0.93)                     | 3.56 (1.06)                   | 3.49 (0.81)                   | 3.48<br>(0.03) | 3.35 (0.97)                    | 3.52 (1.01)                    | 3.49 (0.90)                    | 3.44<br>(0.97) |
| $M_{Negative}$ | 3.29 (0.88)                     | 3.15 (1.03)                   | 3.18 (0.92)                   | 3.22<br>(0.94) | 3.36 (0.92)                     | 3.19 (1.07)                   | 3.34 (1.01)                   | 3.30<br>(1.00) | 3.32 (0.90)                    | 3.17 (1.05)                    | 3.25 (0.97)                    | 3.26<br>(0.97) |
| $M_{Total}$    | 3.30 (0.03)                     | 3.31 (0.04)                   | 3.33 (0.04)                   | 3.10<br>(0.98) | 3.38 (0.03)                     | 3.38 (0.04)                   | 3.42 (0.04)                   | 3.39<br>(0.98) | 3.34 (0.02)                    | 3.35 (0.03)                    | 3.37 (0.03)                    | 3.35<br>(0.98) |



language user was a native speaker of Spanish ( $z = 3.81$ ;  $p < .001$ ; Cohen's  $d = 0.22$ ) or a balanced bilingual ( $z = 4.63$ ;  $p < .001$ ; Cohen's  $d = 0.37$ ) rather than a native speaker of English. A three-way interaction between framing, the language in which the survey was presented and the type of language user showed that when bilingual users were confronted with a Spanish survey, the framing effect was somewhat smaller than the separate two-way interactions suggest ( $z = 2.28$ ;  $p = 0.01$ ; Cohen's  $d = 0.25$ ).

Overall, the sum of the different interactions suggests that there is a framing effect across the board. This effect is larger for, primarily, Spanish speakers who face the survey in Spanish and to a lesser extent also for bilinguals who face the survey in Spanish.

*Survey question polarity: Political attitude statements.* Results (see Table 3) indicate that there are no effects of framing, the language in which the survey was written or the language learner type: the model did not improve significantly by adding any main effect or interaction effect term.

#### 4. Conclusion study 1

Study 1 shows no evidence for a Foreign Language Effect. Framing always had the same effect in L1 and L2 independent of whether a framing effect was overall present (risky-choice framing, attribute framing) or not (framing in surveys), and this was true for both the two groups of unbalanced bilinguals and for balanced bilinguals.<sup>2</sup>

One explanation for the lack of support for an FLE may be that our experimental design was complex, resulting in ample power to detect medium or large effects, but maybe failing to detect smaller ones. Therefore, we ran a second study with a simpler design and more statistical power.

Furthermore, in order to be sure whether we cannot establish an FLE at all, or fail to replicate it for valence framing effects specifically, we also add a moral dilemma to the second experiment. As argued in the introduction, one of the possible explanations of the FLE is a weaker emotional response in L2. However, the framing tasks might be considered a rather "cold" task and hence the moral dilemma could be considered a "hot" task that appeals more to the emotion and value systems (Van Berkum et al., 2009). The evidence regarding the underlying psychological factors driving moral judgements is

**Table 3.** Mean attitude for English L1, Spanish L1 and Balanced Bilinguals (in short: bilinguals) (SD between brackets).

|                | English survey (N=248) |                   |                   |                   | Spanish survey (N=227) |                   |                   |                   | Total              |                    |                    |                    |                    |                    |
|----------------|------------------------|-------------------|-------------------|-------------------|------------------------|-------------------|-------------------|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|                | English L1 (N=103)     |                   | Bilinguals (N=63) |                   | English L1 (N=104)     |                   | Bilinguals (N=59) |                   | English L1 (N=207) |                    | Bilinguals (N=122) |                    | Total              |                    |
|                | Spanish L1 (N=69)      | Spanish L1 (N=77) | Spanish L1 (N=77) | Spanish L1 (N=59) | Spanish L1 (N=77)      | Spanish L1 (N=59) | Spanish L1 (N=77) | Spanish L1 (N=59) | Spanish L1 (N=146) | Spanish L1 (N=146) | Spanish L1 (N=193) | Spanish L1 (N=193) | Spanish L1 (N=146) | Spanish L1 (N=146) |
| $M_{Positive}$ | 3.25 (2.02)            | 3.21 (1.97)       | 3.21 (1.97)       | 3.15 (1.82)       | 3.21 (1.96)            | 3.17 (1.97)       | 3.22 (1.92)       | 3.20 (1.97)       | 3.23 (1.99)        | 3.19 (1.96)        | 3.19 (1.93)        | 3.21 (0.02)        | 3.19 (1.96)        | 3.21 (0.02)        |
| $M_{Negative}$ | 3.17 (2.28)            | 2.96 (1.96)       | 3.14 (1.97)       | 3.14 (1.97)       | 3.18 (2.02)            | 3.27 (1.95)       | 3.29 (1.85)       | 3.24 (1.98)       | 3.17 (2.15)        | 3.13 (1.99)        | 3.21 (1.87)        | 3.17 (2.03)        | 3.13 (1.99)        | 3.17 (2.03)        |
| $M_{Total}$    | 3.21 (2.15)            | 3.09 (1.97)       | 3.15 (1.84)       | 3.15 (1.84)       | 3.19 (1.99)            | 3.22 (1.97)       | 3.26 (1.96)       | 3.22 (1.98)       | 3.20 (2.07)        | 3.16 (1.98)        | 3.20 (1.90)        | 3.19 (2.00)        | 3.16 (1.98)        | 3.19 (2.00)        |

mixed (which we will return to in the general discussion), but the line of reasoning is as follows.

In moral decisions, intuitive processes generally support judgments that favour the essential rights of a person (deontological judgments), while rational processes are supporting judgments favouring the greater good (utilitarian judgements). When classic moral dilemmas like the trolley problem (Thomson, 1976) are presented in an L2, there is some evidence that people are more likely to make utilitarian judgments – as their weaker emotional response in a foreign language leads individuals to be less affected by an emotional aversion to pushing a man onto a track to avoid the death of others. This presumably promotes more utilitarian decisions, which are based on analytical reasoning (e.g. Białek et al., 2019; Cipolletti et al., 2016; Corey et al., 2017; Costa et al., 2014b; Geipel et al., 2015b; Hayakawa et al., 2017).

## 5. Method study 2

In order to increase statistical power, we used a simplified design for Experiment 2. Furthermore, we included another decision task to this second study: a footbridge version of the trolley dilemma. As discussed above, in this task emotional processes would support deontological judgments, whereas more rational processes support utilitarian judgments. Costa et al. (2014b) established an FLE with this task, showing more utilitarian judgments in L2 – likely due to weaker emotional responses. This task serves as an additional test to check whether an FLE is observed using a more emotional dilemma. The study was preregistered at [aspredicted.org](https://aspredicted.org/#17761) (#17761).

### 5.1 Design

In our second study we only recruited L1 speakers of English who learned Spanish as L2. They were assigned to one of two versions of the survey in one of two languages (English or Spanish) in a 2 × 2 between-subjects design.

### 5.2 Participants

Calculations in Gpower indicate that for a Logistic regression with an odds ratio of 1.827 (which is the effect size in Costa et al., 2015), (Pr H0 .55, alpha .05 and a power of .9,  $z$  1.96) at least 501 participants are needed. Between December 12, 2018 and

January 1, 2019, we recruited 585 participants through MTURK. Of these participants, 503 met all quality criteria we formulated in our preregistration: participants had to answer at least 3 out of 4 questions on the Spanish language test correctly, had to report English to be their L1, they should not expose straight lining behaviour (providing same scores to at least 19 out of 20 issue framing questions and/or 7 out of 8 attribute framing dilemmas), or have spent less than 5 min on the survey, and last but not least, respondents could not participate without giving their explicit consent prior to starting answering our questions and they were excluded if they had already participated in our first study.

The 503 eligible respondents were randomly assigned to (a) one of two versions of our survey in (b) one of two language (English or Spanish). Participants in Study 2 (Mean age = 32.48;  $SD$  = 8.92) all learned English from birth and the mean age at which they acquired Spanish was 11.92 ( $SD$  = 8.08).

We checked the randomisation across the two survey versions as well as the two different languages for age ( $t(501) = 0.36$ ;  $p = .72$ ) and ( $t(479.6) = 1.29$ ;  $p = .20$ ), gender ( $\chi^2(2) = 2.25$ ;  $p = .33$  and  $\chi^2(2) = 0.19$ ;  $p = .91$ ) and educational level ( $\chi^2(2) = 2.24$ ;  $p = .33$  and  $\chi^2(2) = 4.60$ ;  $p = .10$ ) and observed no differences between groups.

### 5.3 Materials

Experiment 2 consisted of the same three framing tasks as the ones used in Experiment 1, in the same order. As a fourth task, we included a trolley dilemma, taken from Costa et al. (2014b; see Appendix 7 (online)).

### 5.4 Analysis

Similar to Study 1, we analysed the data of the issue framing tasks and the attribute framing tasks in two cross-classified multilevel models, whereas we ran a loglinear uni-level model for the gain/loss framing task and the footbridge dilemma. See Appendix 8 (online) for details on the models and Appendix 9 (online) for details on the contrasts.

## 6. Results experiment 2

*Risky-choice framing.* Table 4 shows the percentage of choices for the sure option (200 lives saved and 400 lives lost). The choice for the sure option depends

**Table 4.** Proportion of choices for the sure option (*SD* between brackets).

|                              | English survey (L1,<br>N=264) | Spanish survey (L2,<br>N=239) | Total       |
|------------------------------|-------------------------------|-------------------------------|-------------|
| <i>M</i> <sub>Positive</sub> | 0.73 (0.45)                   | 0.69 (0.46)                   | 0.71 (0.45) |
| <i>M</i> <sub>Negative</sub> | 0.35 (0.48)                   | 0.36 (0.48)                   | 0.36 (0.48) |
| <i>M</i> <sub>total</sub>    | 0.53 (0.50)                   | 0.53 (0.50)                   | 0.53 (0.50) |

on the frame in which the choice is presented ( $z = 7.80$ ;  $p < .001$ ;  $SE = 7.32$ ;  $p < .001$ ) the sure option is chosen more often in the gain-frame condition than in the loss-frame condition. The *SE* indicates that this framing effect is more than seven times larger than the random sample differences. The other differences between conditions were not significant and therefore there is no evidence for an FLE.

*Attribute framing.* Results demonstrate that there is a main effect of attribute framing ( $z = 6.74$ ;  $p < .001$ ; Cohen's  $d = 0.24$ ): candidates described with positive frames are evaluated more positively (see Table 5). Adding a main effect of language, as well as the interaction between framing and language, did not significantly improve the model.

*Issue framing.* Results indicate an overall framing effect: attitude objects are evaluated more positively when the question is phrased positively ( $z = 4.00$ ;  $p < .001$ ; Cohen's  $d = 0.05$ ; see Table 6). This very small effect did not depend on the language the survey was presented in, nor were any other differences between conditions significant, so therefore there is no evidence for an FLE.<sup>3</sup>

*Footbridge dilemma.* Table 7 displays the proportion of people that would act utilitarian in the footbridge dilemma, meaning they would push one heavy man on the track to save five others. Results indicate that approximately 40.5% of people would push the man on the train track when the dilemma is presented in English, but when people evaluate the dilemma in L2 (i.e. in Spanish) this percentage goes up to 49.8% ( $z = 2.08$   $p = 0.02$ ;  $SE = 2.21$ ;  $p = .03$ ). The *SE* indicates that, compared to the random sample differences, the difference in utilitarian choices between L2 Spanish and L1 English is two times larger. These results are compatible with Costa et al. (2014b), who

**Table 5.** Mean candidate evaluations (*SD* between brackets).

|                              | English survey (L1,<br>N=264) | Spanish survey (L2,<br>N=239) | Total       |
|------------------------------|-------------------------------|-------------------------------|-------------|
| <i>M</i> <sub>Positive</sub> | 3.41 (1.09)                   | 3.40 (0.99)                   | 3.41 (1.05) |
| <i>M</i> <sub>Negative</sub> | 3.17 (1.03)                   | 3.22 (1.03)                   | 3.20 (1.03) |
| <i>M</i> <sub>total</sub>    | 3.29 (1.08)                   | 3.31 (1.02)                   |             |

**Table 6.** Mean attitude for the different experimental groups (*SD* between brackets).

|                              | English survey (L1,<br>N=264) | Spanish survey<br>(L2, N=239) | Total       |
|------------------------------|-------------------------------|-------------------------------|-------------|
| <i>M</i> <sub>Positive</sub> | 3.29 (2.04)                   | 3.28 (2.00)                   | 3.28 (2.02) |
| <i>M</i> <sub>Negative</sub> | 3.16 (2.07)                   | 3.21 (2.15)                   | 3.19 (2.11) |
| <i>M</i> <sub>total</sub>    | 3.23 (2.06)                   | 3.25 (2.07)                   |             |

report 18% utilitarian choices for L1 speakers and 44% utilitarian choices for L2 speakers in their footbridge dilemma.<sup>4</sup>

## 7. Conclusion study 2

Study 2 was a partial replication of Study 1 with a simpler experimental design and one new task: the footbridge dilemma. This allowed for testing our hypothesis with increased statistical power and for an additional task that showed an FLE in previous studies. Study 2 provided evidence for the existence of all types of framing effects, but only for the footbridge dilemma we found an FLE.

## 8. General discussion

We tested the Foreign Language Effect for three types of valence framing. For risky-choice framing an FLE had already been shown in previous research. We aimed at extending FLE research to two more naturalistic types of valence framing: attribute framing in short informative texts and polarity in attitude questions.

We replicated two out of three common types of framing effects, risky-choice and attribute framing, consistently across English and Spanish texts and across balanced and unbalanced bilinguals. In Study 2 and in a combined analysis (see Appendix 10 (online)), we also observed a (very) small framing effect for question polarity – albeit in an unexpected direction compared to previous studies (e.g. Holleman et al., 2016). Polarity effects in attitude questions have always been shown to be relatively small and dependent of the issues addressed (Holleman, 2000).

Counter to expectations based on previous research (e.g. Costa et al., 2014a; Keysar et al., 2012;

**Table 7.** Proportion of people that would push the man of the track.

|                                 | English survey<br>(L1, N=264) | Spanish survey<br>(L2, N=239) | Total       |
|---------------------------------|-------------------------------|-------------------------------|-------------|
| <i>M</i> <sub>Utilitarian</sub> | 0.41 (0.49)                   | 0.50 (0.50)                   | 0.45 (0.50) |

Thoma & Baum, 2019), we did not find evidence for an FLE in these three framing tasks, as framing effects were always comparable for L1 and L2 users. This lack of an FLE cannot be due to the statistical power in our studies: Experiment 1 had a complex experimental design that had the potential of showing an FLE twice for each framing type, but only had sufficient power to detect statistically medium and large-sized effects. Experiment 2 had a more straightforward design and included enough participants to also detect statistically small effects. Moreover, we conducted a combined analysis of both experiments with even more statistical power, which did not reveal an FLE for any valence framing task either. A lack of statistical power is therefore unlikely to be the cause of the absence of FLEs for these framing types, particularly because we did observe an FLE for the moral dilemma in Study 2.

Another methodological line of reasoning to explain why we did not find any FLE for the valence framing tasks could be in the use of a crowdsourcing platform to conduct our experiments. We aimed to obtain a large and diverse group of L2 and L1 users by testing participants through MTurk. Can this web-based platform be trusted as a valid source for language processing data? Other research claims that it can indeed (Gibson et al., 2011; Sprouse, 2011). We included a stern check on participants' L2 proficiency, and cleaned our data rigorously by checking straight-lining behaviour and outliers in response times. Furthermore, the results in Experiment 1 and Experiment 2 are comparable, which adds to the credibility of the data. Moreover, as we did establish the hypothesised framing effects in both of our studies, and an FLE for the footbridge dilemma the explanation that our results are caused by poor data quality is unlikely.

Then why did our experiments not establish an FLE for the risky-choice dilemma, whereas previous studies did? It may be due to characteristics of our participants. Previous studies by Keysar et al. (2012) and Costa et al. (2014a, 2014b) as well as Thoma and Baum (2019) have used student participant groups relatively homogenous in their L2 use and learning background, relying on people who learned their L2 in adolescence or young adulthood (mean age at which participants began learning their L2 ranged from 12–17 across these three studies). In the current study, a more heterogeneous group of L2 users was included (with a mean age of 32–35, and a mean age of acquiring Spanish as a second

language of about 11–14 years, and a mean age of acquiring English as a second language of about 8 years old. In most studies, participants are predominantly students (mean age about 20–22), so probably with a highest level of education of (some) college. In both our studies about 30–40% of participants reported their highest level of education to be “some college”, whereas about 40% had obtained a BA and 12–20% had obtained their MA). The heterogeneity of our participant group might have resulted in a larger variation in responses and hence a reduction of the chance of observing an FLE.

These differences between our study and earlier research may explain the divergent results in two different ways. For one, the higher heterogeneity and presumably higher variations in responses may have increased the error variance and therefore the chance of observing an effect. Second, the fact that our participants have more advanced L2 skills and a more natural use of L2 in daily life in our experiments as compared to earlier studies may count as an explanation. Recent research by Thoma and Baum suggests that FLEs for risky-choice framing tasks disappear once the cognitive load for using and responding in an L2 is not as heavy, either due to a simpler task (see Winkler et al., 2016), or because of a higher L2 proficiency. Similar, recent studies did not find an FLE (Brouwer, 2019; Čavar & Tytus, 2018; Muda et al., 2020 (but see Białek & Fugelsang, 2019, for a challenge of some of the conclusions)) in moral decision making tasks with highly proficient and acculturated L2 speakers.

In contrast, however, we did find an FLE for the trolley dilemma in Study 2 for our English L1 participants. This suggests that there might be an alternative explanation for the non-occurrence of FLEs in the risky-choice and other framing tasks. It might be worthwhile for future research to distinguish between different kinds of emotions. Research suggests that specifically people's socio-cultural norms might be less activated when encountering information in a foreign language than when encountering the same information in their native tongue (Geipel et al., 2015a, 2015b). The idea is that people experience norms in a context where the native language is used and therefore these norms are connected to language-dependent memory. This could explain why we found an FLE for our moral dilemma, and not for our framing tasks, assuming that attribute frames and risky-choice framing tasks evoke attitudinal evaluations, but probably not

necessarily socio-cultural embedded norms. It does not explain, however, why Keysar et al. (2012) and Costa et al. (2014a) did find FLEs for the risky choice-framing tasks in their experiments.

Maybe the fact that we did not find FLEs for our framing effects, in contrast to the results of Keysar et al. (2012) and Costa et al. (2014a), whereas we did find an FLE for our trolley dilemma, can be explained because perhaps the trolley dilemma calls upon a stronger emotional involvement compared to framing tasks, so that this task could show an FLE even with a more proficient group of L2 users, whereas an FLE in risky-choice framing tasks will only emerge with a more homogeneous and less culturally grounded group of L2 users than the participants in our study.

But how does this relate to the fact that our participants did rate their L2 as less emotional? Unlike Keysar et al. (2012) and Costa et al. (2014a), we directly measured the extent of perceived L1-L2 emotionality differences. Our participants did rate their L2 as being less emotional than their L1, as expected (Section 2.2). Nevertheless, the risky-choice framing effect and the overall attribute framing effect were present in both L1 and L2 for all participant groups and we did not establish any FLE for the valence framing tasks, which either demonstrates that these effects can arise even in a state of “disembodied cognition” (Pavlenko, 2012) when people are reading in a foreign language with low(er) emotional grounding, or points to the explanation that processing emotionality is not hampered in participants with a higher L2 proficiency. It may also be the case that the participants’ introspective judgements as to the emotionality differences between their L1 and L2 only relate to their actual language processing when very explicit emotional decision tasks are at stake, such as a moral choice dilemma. The latter also implies that the previous findings by Keysar et al. (2012) and Costa et al. (2014a, 2014b) perhaps need to be reinterpreted in the light of reduced cognitive processing as a possible mechanism of foreign language effects.

It is worthwhile to combine the literature on FLEs in moral dilemmas and framing tasks to further study the mechanisms at play. In the moral dilemma literature there are several theories about underlying psychological factors that drive the judgment, with mixed evidence. Apart from the weaker emotions account causing an FLE, another possible driver could be an increased cognitive load, as shown by Geipel et al. (2016). This factor was also suggested

by Thoma and Baum (2019) to be relevant in framing tasks. Another possible factor driving foreign language effects in moral judgements is vividness of mental imagery (Hayakawa & Keysar, 2018). In a foreign language, mental images are less vivid, possibly partly inhibiting visualisation of the situation. This mechanism could be relevant in framing tasks as well. Maybe framing tasks are in general less susceptible to vivid mental imagery, leaving little opportunity for a reduction in visualisation and hence causing an FLE.

*Concluding.* Current research has not shown evidence for an FLE for any of the three types of valence framing. According to Thoma and Baum (2019) the FLE arises because in L2 an increased cognitive demand results in reduced automatic semantic access and hence in weaker emotions. In our study, participants report a subjectively lower emotionality in the L2 than in the L1, yet they may be relatively highly proficient yielding ample of cognitive resources for emotional lexical access in the three framing tasks that require lower emotional involvement. However, for the more emotionally demanding moral decision task a weaker emotional reaction might still have caused the observed FLE. Of course future research still needs to determine whether the framing tasks really does require lower emotional involvement than the moral decision task.

*Directions for future research.* First, recent research (Brouwer, 2019) tested whether the FLE holds for higher language proficiencies and corroborates this idea for two quite closely related languages (English-Dutch). It would be interesting to explore the FLE for advanced unbalanced highly proficient L2 users further, for different language combinations. This may also point to a more cognitive explanation for the existence or lack of an FLE: with highly proficient users, the cognitive task load may become comparable to that of L1 users, as Thoma and Baum (2019) suggest.

Second, maybe the relationship between L2 contexts and framing effects will only arise when hardly any emotions resonate, that is, when L2 has been learned (exclusively) in a classroom context. In contrast, the FLE would disappear if the L2 was acquired in a more emotional natural language context. This hypothesis, related to emotional embodiment, can be tested by distinguishing proficient L2 users who acquired their L2 exclusively in a classroom context versus proficient L2 users who acquired their L2 predominantly in real-life situations.



Third, it is worthwhile to explore the intriguing mismatch between participants' own judgement of emotionality of their L2 versus the actual (non)appearance of an FLE: an FLE does arise for the moral decision task, but not for the valence framing effects. It might be that emotions have been defined too one-dimensionally in this FLE research, and we should distinguish between different kinds of emotions, such as socio-cultural norms, attitudes, and feelings. It may also be the case that emotions do play a role in valence framing effects, but these emotions are equally accessible for L1 users and proficient and/or acculturated L2 users. In that case, maybe the L1/L2 paradigm is not a subtle enough measure of emotional resonance, and other measures (e.g. Van Berkum, 2018) should be obtained to show the extent of emotionality in processing valence frames, such as physiological measures like skin responses.

## Notes

1. We report the results based on  $N = 475$  participants as these analyses have also been reported in conference proceedings of the ICA 2018 conference (Prague) and to create maximum statistical power for our test. We also ran the same analyses applying stricter quality criteria (identical to those in study 2, preregistered in aspredicted.org #17761), removing people who took less than 5 min to answer the survey questions ( $N = 1$ ) and straight liners on at least one of the tasks ( $N = 14$ ). This did not alter the conclusions.
2. We found larger framing effects in Spanish materials and for Spanish L1 users. As our focus is on FLE, and we did not hypothesize this effect, we do not discuss it in detail here. It is tempting to attribute this effect to cultural priming, i.e. a supposedly higher emotionality of the Spanish language or native speakers of Spanish (compared to English or native speakers of English), but we are not aware of existent research to support this suggestion.
3. In order to maximize the power and chance of finding a Foreign Language effect we also ran a combined analysis of Study 1 and 2 (for L1 English speakers) for all three types of framing. This analysis gave similar results compared to Study 2 (see Appendix 10 (online) for details).
4. To check whether our footbridge dilemma scores are not the result of random answering behavior, we conducted a one sample  $t$ -test. The percentage of utilitarian choices in English indeed differs from chance ( $t = 3.13$ ,  $df = 263$ ,  $p = .002$ ) as does the overall percentage of utilitarian choices ( $t = 2.28$ ,  $df = 502$ ,  $p = .02$ ). Hence, it seems unlikely that our results are caused by random responding.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Bregje Holleman  <http://orcid.org/0000-0003-2208-8018>  
 Naomi Kamoen  <http://orcid.org/0000-0002-8719-2417>  
 Marijn Struiksmā  <http://orcid.org/0000-0002-1166-1424>

## References

- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. doi:10.1037/1089-2680.5.4.323
- Bialek, M., & Fugelsang, J. (2019). No evidence for decreased foreign language effect in highly proficient and acculturated bilinguals: A commentary on Čavar and Tytus (2018). *Journal of Multilingual and Multicultural Development*, 40(8), 679–686. doi:10.1080/01434632.2018.1547072
- Bialek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, 85, 103855. doi:10.1016/j.jesp.2019.103855
- Brouwer, S. (2019). The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *Journal of Multilingual and Multicultural Development*, 40(10), 865–878. doi:10.1080/01434632.2019.1585863
- Čavar, F., & Tytus, A. E. (2018). Moral judgement and foreign language effect: When the foreign language becomes the second language. *Journal of Multilingual and Multicultural Development*, 39(1), 17–28. doi:10.1080/01434632.2017.1304397
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, 29(1), 23–40. doi:10.1080/09515089.2014.993063
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Colbeck, K. L., & Bowers, J. S. (2012). Blinded by taboo words in L1 but not L2. *Emotion*, 12(2), 217–222. doi:10.1037/a0026387
- Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1109. doi:10.1037/xlm0000356
- Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014a). “Piensa” twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254. doi:10.1016/j.cognition.2013.11.010
- Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2015). Corrigendum to “piensa” twice: On the foreign language effect in decision making. *Cognition*, 142, 362–363. doi:10.1016/j.cognition.2015.05.013
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014b). Your morals depend on language. *PLoS ONE*, 9(4), e94842. doi:10.1371/journal.pone.0094842
- Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition*, 15(1), 181–189. doi:10.1017/S1366728911000095
- De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, 313(5787), 684–687. doi:10.1126/science.1128356



- Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the Speech of multilinguals. *Journal of Multilingual and Multicultural Development*, 25(2–3), 204–222. doi:10.1080/01434630408666529
- Dewaele, J.-M., & Pavlenko, A. (2001–2003). *Web questionnaire on bilingualism and emotion* [Unpublished manuscript]. University of London. [https://www.academia.edu/2134800/Web\\_questionnaire\\_bilingualism\\_and\\_emotions](https://www.academia.edu/2134800/Web_questionnaire_bilingualism_and_emotions).
- Eilola, T. M., & Havelka, J. (2010). Behavioural and physiological responses to emotional and taboo stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, 15(3), 353–369. doi:10.1177/1367006910379263
- Election Compass. (2016). <http://usa.electioncompass.org/>.
- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion*, 8(4), 453–457. doi:10.1037/a0012887
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15(2), 115–141. doi:10.1080/02699930125908
- Gao, S., Zika, O., Rogers, R. D., & Thierry, G. (2015). Second language feedback abolishes the “hot hand” effect during even-probability gambling. *Journal of Neuroscience*, 35(15), 5983–5989. doi:10.1523/JNEUROSCI.3622-14.2015
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, 59, 8–17. doi:10.1016/j.jesp.2015.02.001
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). The foreign language effect on moral judgment: The role of emotions and norms. *PLOS ONE*, 10(7), e0131529. doi:10.1371/journal.pone.0131529
- Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition*, 154, 34–39. doi:10.1016/j.cognition.2016.05.010
- Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, 5(8), 509–524. doi:10.1111/j.1749-818X.2011.00295.x
- Hadjichristidis, C., Geipel, J., & Surian, L. (2017). How foreign language affects decisions: Rethinking the brain-drain model. *Journal of International Business Studies*, 48(5), 645–651. doi:10.1057/s41267-016-0040-1
- Harris, C., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression and representation* (pp. 257–283). Clevedon: Multilingual Matters Ltd.
- Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24(4), 561–579. doi:10.1017/S0142716403000286
- Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition*, 173, 8–15. doi:10.1016/j.cognition.2017.12.010
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, 28(10), 1387–1397. doi:10.1177/0956797617720944
- Holleman, B. (2000). *The forbid/allow asymmetry. On the cognitive mechanisms underlying wording effects in surveys. Utrecht studies in language and communication*. Rodopi.
- Holleman, B., Kamoen, N., Krouwel, A., van de Pol, J., & de Vreese, C. (2016). Positive vs. Negative: The impact of question polarity in voting advice applications. *PLoS ONE*, 11(10), e0164184. doi:10.1371/journal.pone.0164184
- Holleman, B. C., & Maat, H. L. P. (2009). The pragmatics of profiling: Framing effects in text interpretation and text production. *Journal of Pragmatics*, 41(11), 2204–2221. doi:10.1016/j.pragma.2008.09.038
- Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PLoS ONE*, 12(10), e0186027. doi:10.1371/journal.pone.0186027
- ISideWith. (2016). [isidewith.com/polls](http://isidewith.com/polls).
- Jończyk, R., Boutonnet, B., Musiał, K., Hoemann, K., & Thierry, G. (2016). The bilingual brain turns a blind eye to negative statements in the second language. *Cognitive, Affective & Behavioral Neuroscience*, 16(3), 527–540. doi:10.3758/s13415-016-0411-x
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. doi:10.1037/0003-066X.58.9.697
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. doi:10.2307/1914185
- Kamoen, N., Holleman, B., & van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181–189.
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668. Clevedon, England: Multilingual Matters doi:10.1177/0956797611432178
- Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The Psychology of evaluation: Affective processes in cognition and emotion* (pp. 9–50). Lawrence Erlbaum Associates.
- Kühberger, A. (1998). The influence of framing on risky decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes*, 75(1), 23–55. doi:10.1006/obhd.1998.2781
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15(1), 374–378. doi:10.1086/209174
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149–188. doi:10.1006/obhd.1998.2804
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267–286. doi:10.1037/0033-2909.127.2.267
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. doi:10.1044/1092-4388(2007)067
- Muda, R., Pieńkosz, D., Francis, K. B., & Białek, M. (2020). The moral foreign language effect is stable across presentation modalities. *Quarterly Journal of Experimental Psychology*, 73(11), 1930–1938. doi:10.1177/1747021820935072

- National Spanish Examinations. (2015). nationalspanishexam.org.
- Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961–1967. doi:10.1016/j.neuropsychologia.2012.04.021
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6), 405–428. doi:10.1080/00207594.2012.743665
- Quené, H., & van den Bergh, H. (2004). On multilevel modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1-2), 103–121. doi:10.1016/j.specom.2004.02.004
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425. doi:10.1016/j.jml.2008.02.002
- Rozin, P., Berman, R., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, 24(3), 536–548. doi:10.1080/02699930902793462
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320. doi:10.1207/S15327957PSPR0504\_2
- Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly*, 5(1), 91–92. doi:10.1086/265467
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. Sage Publications.
- Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43(1), 155–167. doi:10.3758/s13428-010-0039-7
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1), 67–85. doi:10.1037/0033-2909.110.1.67
- Thoma, D., & Baum, A. (2019). Reduced language processing automaticity induces weaker emotions in bilinguals regardless of learning context. *Emotion*, 19(6), 1023–1034. doi:10.1037/emo0000502
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204–217. doi:10.5840/monist197659224
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4), S251–S278. doi:10.1086/296365
- Van Berkum, J. J. A. (2018). *Language comprehension, emotion, and sociality: Aren't we missing something?* In S. A. Rueschemeyer & G. Gaskell (Eds.), *Oxford handbook of psycholinguistics* (pp. 644–669). Oxford University Press.
- Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong?: The brain's fast response to morally objectionable statements. *Psychological Science*, 20(9), 1092–1099. doi:10.1111/j.1467-9280.2009.02411.x
- Van den Bergh, H. (1990). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement*, 14(1), 1–12. doi:10.1177/014662169001400101
- Winkielman, P., Knutson, B., Paulus, M., & Trujillo, J. L. (2007). Affective influence on judgments and decisions: Moving towards core mechanisms. *Review of General Psychology*, 11(2), 179–192. doi:10.1037/1089-2680.11.2.179
- Winkel, H., Ratitamkul, T., Brambley, V., Nagarachinda, T., & Tiencharoen, S. (2016). Decision-making and the framing effect in a foreign and native language. *Journal of Cognitive Psychology*, 28(4), 427–436. doi:10.1080/20445911.2016.1139583