

**The moral foreign language effect is stable across presentation modalities**

*Rafał Muda<sup>1,2</sup>, Damian Pieńkosz<sup>1</sup>, Kathryn B. Francis<sup>3</sup> & Michał Białek<sup>4</sup>*

*1 Faculty of Economics, Maria Curie-Skłodowska University, Poland*

*2 Institute of Psychology, Maria Curie-Skłodowska University, Poland*

*3 Division of Psychology, University of Bradford, UK*

*4 Institute of Psychology, University of Wrocław, Poland*

*Correspondence to:*

Michał Białek, PhD;

Institute of Psychology, University of Wrocław, Dawida 1, 50-52 Wrocław, Poland; email:  
mihalbialek@gmail.com

**Word counts**

Main text: 4289

This is the accepted version of the manuscript (following peer review).  
This article has been accepted for publication in the Quarterly Journal  
of Experimental Psychology.

### **Abstract**

Peoples' judgments and decisions often change when made in their foreign language. Existing research testing this foreign language effect has predominantly used text-based stimuli with little research focusing on the impact of *listening* to audio stimuli on the effect. The only existing study on this topic found shifts in people's moral decisions only in the audio modality. Firstly, by reanalyzing the data from this previous study and by collecting data in an additional experiment, we found no consistent effects of using foreign language on moral judgments. Secondly, in both datasets we found no significant language by modality interaction. Overall, our results highlight the need for more robust testing of the foreign language effect, and its boundary conditions. However, modality of presentation does not appear to be a candidate for explaining its variability. Data and materials for this experiment are available at <https://osf.io/qbjxn/>.

**Keywords:** foreign language effect, bilingualism, moral judgments, modalities

## **The moral foreign language effect is stable across presentation modalities**

### **Introduction**

Many decisions are made in a foreign language from those of great importance such as negotiating and communicating in international organizations to everyday decisions such as those made by expats that have just arrived in a new country. There are far-reaching consequences of this process: using a foreign language prompts people to make different, and usually better decisions, compared to when using their native language (the so called “foreign language effect”; Keysar, Hayakawa, & An, 2012). The foreign language effect has been documented in many domains of judgment and decision making. Among other benefits, people using a foreign language are less risk averse and more willing to take risks with positive expected value (e.g. Costa, Foucart, Arnon, Aparici, & Apesteguia, 2014, Keysar, Hayakawa, & An, 2012, but see Hayakawa, Lau, Holtzmann, Costa, & Keysar, 2019 for boundary conditions), and are less susceptible to illusory correlations (Gao, Zika, Rogers, & Thierry, 2015) and to superstitious beliefs (Hadjichristidis, Geipel, & Surian, 2017). Foreign language use also affects our judgments: it makes people perceive risks as lower and benefits as higher (Hadjichristidis, Geipel, & Savadori, 2015) and it influences moral judgments (e.g. Bialek, Paruzel-Czachura, & Gawronski, 2019b; Costa, et al., 2014a; Geipel, Hadjichristidis, & Surian, 2015a, 2015b; Muda, et al., 2018).

To date, almost all previous studies have tested the foreign language effect using text stimuli. However, many decisions in a foreign language are made when *listening* to a foreign language instead of reading it and there is still a lack of systematic evidence showing whether the foreign language effect works in this more ecologically valid and broadly occurring circumstance. Specifically, from an evolutionary perspective, people have always communicated verbally whereas writing has been a relatively new means of communication (Kroll, 1981). Developmentally, children first communicate verbally and are only later taught to read and write (Libermann, 1992). Finally, the majority of our informal and everyday interactions happen verbally (although, large parts of these, now happen online using text, Valkenburg, & Peter, 2011). Considering all the above, the audio modality requires extra attention from researchers investigating the foreign language effect.

To this point however, the only existing research comparing foreign language effect in verbal vs. text stimuli investigated moral decision making and found the foreign language effect in the audio modality only (Brouwer, 2019). Past research, however, has also successfully documented the foreign language effect with written vignettes. Hence, the aim of the current study is to test whether the foreign language effect differs when individuals make judgments on the basis of auditory stimuli compared to text-based stimuli.

We focus our research question regarding modality to moral decision making because of its overreaching character and importance for practice (Goodwin, Piazza, & Rosin, 2014) and because the foreign language effect has been extensively documented in the moral domain. The moral foreign language effect refers to changes in moral judgments, typically reported to be more utilitarian. For example, when using their foreign language, people assume harm directly caused to others, but which saves more from similar harm, to be more acceptable (Costa, et al., 2014a; Geipel, Hadjichristidis, & Surian, 2015a, 2015b; Nakamura, 2015). The moral foreign language effect has been explained by a few potential mechanisms. Specifically, some authors claim that using a foreign language (1) reduces activation of social and moral norms (Gawinkowska, Paradowski, & Bilewicz, 2013; Geipel, Hadjichristidis, & Surian, 2015b); (2) reduces the importance of the consequences of an action (Białek, et al., 2019b), (3) reduces the strength of deontological and utilitarian inclinations (Muda et al., 2018), (4) shifts the attention from intentions to outcomes when judging other's moral actions (Geipel, Hadjichristidis, & Surian, 2016), and (5) increases psychological distance and reduces emotional responses (Costa et al., 2014a; Corey et al., 2017). Regardless of which mechanism affects moral judgments, these changes affect moral inclinations *per se*, rather than broad decision-making strategies such as decision avoidance or inaction aversion (Białek, et al., 2019b). All of these effects are explained in the dual-process terminology by either claiming that the intuitive input is weakened (e.g., because foreign language produces weaker affective responses, Caldwell-Harris, 2015), that people are more likely to engage in reflection (because processing foreign language requires greater effort and activates more cognitive resources, (Hayakawa, Costa, Foucart, & Keysar, 2016), or that people are less sensitive to the conflict between several simultaneously processed intuitions (Bialek, Muda, Stewart, Niszczoła, Pieńkosz, 2019).

Despite having substantial knowledge about the potential mechanisms behind the moral foreign language effect, we still do not know whether it works in situations that closely mimic real-life (i.e. when listening to a foreign language instead of reading in it). To our best knowledge, only one study has tried to test this on a group of highly proficient foreign language speakers, showing that listening to moral dilemmas in a foreign language results in increased acceptance of utilitarian outcomes compared to the native language, whereas reading the same moral dilemmas in different languages produces no effect (Brouwer, 2019). This is a noteworthy finding, suggesting that the foreign language effect is amplified when discussing a problem orally. However, such a strong claim requires strong evidence, beyond that of a single empirical investigation (Zwaan et al, 2018). There are further reasons why it is worth investigating this topic: we have identified methodological issues in Brouwer (2019). Firstly, the author did not provide a statistical test demonstrating that the foreign language effect differs between text and audio conditions. More precisely, in Experiment 1, the

author tested whether the foreign language effect exists in a text condition observing no impact of foreign language use and separately, in Experiment 2, tested whether the foreign language existed in an audio condition, observing a significant impact of foreign language use. However, no direct test for different-sized effects was conducted. As Wickelmaier (2015) noted regarding a similar issue: *“the difference between significant and not significant is not itself significant (Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011). Rather, a test of the difference of the effects is required”* (p. 1). Secondly, the tested sample sizes were small ( $n = 60$  in both experiments), giving  $\beta = .8$  power to detect only improbably large effect sizes of  $d = .64$ . Power to detect a more plausible estimate of the size of the foreign language effect (at  $d = .3$ , Białek & Fugelsang, 2019) was much lower,  $\beta = .31$ . To better illustrate the power issue here, with the sample sizes used in Brouwer (2019) there is only a 31% chance to observe a true effect of a moderate effect size (Funder & Ozer, 2019), and only about 9% chance to find it twice in a row. Finding or not finding an effect in such small samples is therefore attributable to luck, rather than to the existence or non-existence of a true effect in a population (Simonsohn, 2015). Even more luck would be required to detect a difference between the effects in two samples, since it requires at least four times the sample required for detection of the main effect (Goodhue, Lewis, & Thompson, 2007). Indirect evidence of low power status of the experiment is the failure to find a well-established effect whereby people tend to approve of utilitarian actions in impersonal dilemmas more than in personal ones.<sup>1</sup> Typically, people permit redirecting a trolley more than they permit pushing a large person on some tracks, both in order to save five people, (e.g. Geipel, Hadjichristidis, & Surian, 2015a; Moore et al., 2011; Paharia et al., 2009; Roysman & Baron, 2002). This preference for indirect harm has not been observed in Brouwer (2019).

We addressed these issues in two stages. In Stage 1, we reanalyzed raw data from Brouwer (2019) by collapsing the data from the two experiments and increasing the power. In Stage 2, we also conducted a preregistered experiment directly testing whether differences between text and audio foreign language effects exist.

### **Stage 1: Reanalysis of Brouwer (2019)**

---

<sup>1</sup> Impersonal moral dilemmas are those in which harm is caused indirectly and could be seen as a side-effect of saving others, e.g. to hit a switch that will direct the trolley onto the alternate track where it will kill just one person instead of five. Personal moral dilemmas are those in which harm is caused by active and direct act and can be seen as a means of saving others, e.g. to push the stranger from the footbridge to prevent the oncoming trolley from killing five people.

Thanks to Dr Susanne Brouwer, we were able to access the raw data from experiments 1 and 2 published in Brouwer (2019). Given procedural similarities<sup>2</sup>, we collapsed the data from the two experiments, testing for differences in the foreign language effect across modalities (i.e., text-based and audio-based vignettes). First, we summarized affirmative responses to personal (Footbridge, Crying baby, Vitamins) and impersonal scenarios (Switch, Lost wallet, Taxes) separately. This allowed us to create two 0-3 utilitarianism indexes for each type of dilemma. This allowed us to analyze decisions as a continuous variable. We submitted the indexes to a 2 (dilemma type: personal, impersonal; within-subject) x 2 (language: native, foreign; between-subject) x 2 (modality: text, audio; between-subjects) mixed model ANOVA, and to its Bayesian version, using JASP 0.11.1 (Love, et al., 2019). While traditional, p-value based inference does not allow to support the null, but only to reject it or not, Bayesian analysis (Bayes Factor, B) allows analyses to support the null effect (Dienes, 2014). The notation  $B_{10}$  suggest evidence for the hypothesis against the null, and  $B_{01}$  suggests the evidence is for the null against the tested effect, e.g.,  $B_{01} = 6$  means that the data is six times more likely to be observed under null than under the tested model. Although allowing for continuous interpretation,  $B < 3$  provides only weak evidence for any model and is considered inconclusive, and  $B < 10$  as providing only moderate evidence. (Lee & Wagenmakers, 2013).

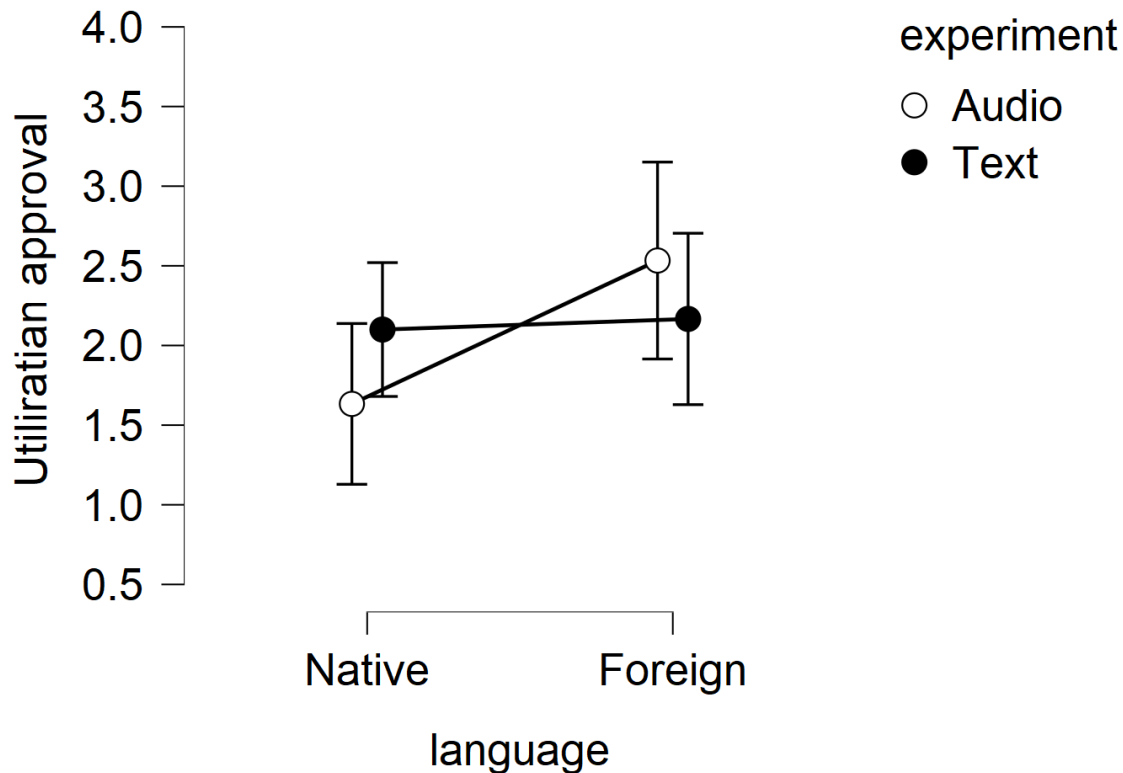
For the first analysis, we used data from all 120 participants, not excluding any based on their proficiency (similarly as done in the original paper). Figure 1 presents the results of the experiment. We found a significant effect of dilemma type,  $F(1, 116) = 5.51$ ,  $p = .021$ ,  $\eta_p^2 = .05$ ,  $B_{01} = 1.24$ , in which participants found it more permissible to act in impersonal ( $M = 1.16$ ,  $CI_{95\%}[1.00-1.31]$ ) moral dilemmas compared to personal moral dilemmas ( $M = 0.93$ ,  $CI_{95\%}[0.80-1.10]$ ). However, this analysis is inconclusive in Bayesian terms. Although non-significant, the main effect of language was close to the cut-off value of significance,  $F(1, 116) = 3.55$ ,  $p = .062$ ,  $\eta_p^2 = .03$ ,  $B_{01} = 2.31$ , with greater permissibility of the sacrifice in foreign language ( $M = 1.18$ ,  $CI_{95\%}[1.00-1.36]$ ) compared to native language ( $M = 0.93$ ,  $CI_{95\%}[0.75-1.11]$ ). Also this analysis is inconclusive in Bayesian terms. Our interest was in the modality by language interaction, which happened to be non-significant,  $F(1, 116) = 2.64$ ,  $p = .107$ ,  $\eta_p^2 = .022$ ,  $B_{01} = 6.15$ , with only moderate evidence against it in Bayesian terms. All other effects were non-significant,  $p > .107$ ,  $B_{01} > 7$ . All of these analyses suggest the available evidence is relatively weak, and does not allow us to draw strong conclusions either way.

*Figure 1*

---

<sup>2</sup> The two experiments had some procedural differences, (e.g. timing of presentation, ability to review the presentation) that could have their own effects on the moral responses. These are, however, likely small and should not interact with the foreign language effect, allowing us to validly collapse the experiments. Otherwise, almost no meta-analysis would be justified.

Utilitarian choices depending on language and mode of presentation in collapsed data of Brouwer (2019).



*Note.* Utilitarian approval is a sum of affirmative responses to six personal and impersonal dilemmas. Error bars represent the 95% credible intervals.

To be consistent with typical procedure in the foreign language effect research, we also repeated the analysis excluding participants with low self-reported proficiency (calculated as a mean of four self-reported dimensions of proficiency: reading, writing, speaking, and oral comprehension). We retained participants who self-rated their proficiency to be on average 3 in a scale 1-5, or higher ( $n = 103$ ). We did so, because results of the experiment could have been contaminated by participants with poor comprehension of the vignettes. These participants could have simply responded randomly. Increased noise in the data caused by such participants would make it hard to detect even the most robust effects in moral psychology, or could create random effects. After reducing the sample size, effects of language and language  $\times$  modality interaction remained non-significant ( $p = 0.089$  and  $p = 0.220$  respectively) and even the dilemma type effect turned out to be non-significant ( $p = .072$ ). This result suggests that the failure to detect the modality effects on moral foreign language was not due random responding of low proficient participants, but rather due to its non-existence or to low experimental power.

Owing to increased power by combining the datasets, we were able to verify previous findings that direct action in personal dilemmas is less permissible than indirect action in impersonal dilemmas. Most of the test statistics were found to be inconclusive. Yet, the collected data did not support the claim for attenuating effects of textual presentation of the stimuli in the foreign language. To further explore this issue, we collected an independent set of data, using different vignettes.

### **Stage 2: Experimental Investigation of Foreign Language Effect in Text versus Audio**

#### **Method**

The experiment was preregistered: <http://aspredicted.org/blind.php?x=22jg6u>. Data and materials for this experiment is available at <https://osf.io/qbjxn/>.

#### ***Participants and Design***

We initially recruited 186 participants but excluded data from 21 individuals who failed the translation task<sup>3</sup>, thus we analyzed data from  $N = 165$  participants (89 females,  $M_{\text{Age}} = 22.70$ ,  $SD = 3.99$ ) who were recruited through an external research firm and paid for their time (a flat fee of 20 PLN or approximately \$5). All participants met our preregistered selection criteria (which also follows widely adopted selection criteria by Keysar, Hayakawa, & An, 2012 and Costa et al., 2014a, Muda, et al., 2018): stated that Polish is their native language, did not have a parent that speaks English as their native language, did not live more than 10 months in an English speaking country, self-rated their global proficiency in English as 5 or higher on a scale from 1 to 10. The mean acquisition age of English was at the age 7.87 years ( $SD = 2.21$ ). Details about participants' proficiency can be found in Table 1.

Participants were randomly assigned (true, not quota-based randomization) to both one language condition and one presentation mode condition in a  $2 \times 2$  between-subjects design (see Table 1 for details). The experiment took approximately 10 minutes.

---

<sup>3</sup> We were interested in testing the  $2$  (language)  $\times$   $2$  (modality) interaction with between-subjects ANOVA. To this end, we determined the sample size based on a-priori power analysis using G\*power (Faul, et al., 2007): power = 0.95; effect size  $f = 0.30$  (based on previous research on moral foreign language effect, for example see Geipel, Hadjichristidis, & Surian, 2015a,b);  $p = 0.05$ ; number of groups = 4 (language: native vs. foreign; modality: text vs. audio), numerator  $df = 1$ . The analysis indicated a sample size of 147 participants. We collected more data due to planned exclusions (i.e. failure in translation test), but slightly fewer than preregistered  $n = 200$  because of no-shows at the very end of our lab reservation.



Table 1

Proficiency data by experimental design with condition assignment

| Language<br>Presentation Mode       | Native (Polish) |              | Foreign (English) |              |
|-------------------------------------|-----------------|--------------|-------------------|--------------|
|                                     | Text            | Audio        | Text              | Audio        |
| <b>N</b>                            | 39              | 62           | 28                | 36           |
| <b>Female (in per cent)</b>         | 54              | 48           | 50                | 31           |
| <b>Age</b>                          | 23.28 (6.82)    | 22.56 (2.85) | 22.21 (1.85)      | 22.67 (2.58) |
| <b>Foreign language proficiency</b> | 7.26 (1.29)     | 7.23 (1.47)  | 7.32 (1.22)       | 7.25 (1.27)  |
| <b>Age of acquisition</b>           | 7.69 (2.47)     | 8.21 (2.14)  | 7.43 (1.73)       | 7.81 (2.34)  |
| <b>Reading</b>                      | 7.79 (1.54)     | 7.50 (1.37)  | 7.71 (1.12)       | 7.75 (1.27)  |
| <b>Writing</b>                      | 6.97 (1.61)     | 6.68 (1.62)  | 6.82 (1.47)       | 7.06 (1.59)  |
| <b>Speaking</b>                     | 6.51 (2.05)     | 6.63 (1.70)  | 6.82 (1.57)       | 6.86 (1.90)  |
| <b>Oral comprehension</b>           | 7.59 (1.60)     | 7.11 (1.77)  | 7.57 (1.71)       | 7.53 (1.54)  |

**Note:** Data presented as mean (SD). All proficiency measures were tested using 10-point Likert scale. Participants with understanding below 5 were excluded from the analysis, and are not presented here.

### Materials and Procedure

To familiarize participants with the experimental task, before the experimental (moral) scenarios were presented, each participant received two training (non-moral) dilemmas (Train and Headache dilemmas, for the full text see the <https://osf.io/qbjxn/>) framed in the same language and modality as in the subsequent experimental scenarios. These training sessions were important as lack of familiarity with the procedure, especially for participants less-versed in computer use and technology, could influence responding. For example, novelty effects could have increased cognitive engagement (Spears et al., 2018), or imposed cognitive load (Białek & De Neys, 2017), both of which are suggested to affect moral decision making.

After the training round, participants responded to three moral dilemmas presented to them in randomized order: Footbridge, (Costa et al., 2014b); Terrorist, (Corey et al., 2017), and Lost Wallet (Geipel, Hadjichristidis, & Surian, 2015b)<sup>4</sup>. All scenarios have been used in previous research on the foreign language effect, showing higher utilitarian tendency when making judgments in a foreign language. Participants were asked to make an appropriateness judgment ('Is it appropriate for you to [specific to scenario]?') and could respond on a Likert scale (ranging from 1: extremely inappropriate

<sup>4</sup> Two out of three dilemmas were also used in Brouwer (2019).

to 7: extremely appropriate). Participants responded in this way in both the text and audio modality conditions.

The audio version of dilemmas was created with use of Amazon Polly software “*that turns text into lifelike speech*” (<https://aws.amazon.com/polly/>). The voices used in specific language versions were Kendra (in English) and Ewa (in Polish). We used Amazon Polly (a text-to-speech system) to avoid potential problems with accented speech: such speech is not only more difficult to comprehend (Crowther, Trofimovich, & Isaacs, 2016), but an accented speaker is also judged as less credible (Lev-Ari & Keysar, 2010). Previous research utilizing Amazon Polly observed promising results in terms of use Amazon Polly in behavioral research, suggesting that artificial voices sound quite naturally (Jeong, Lee, & Kang, 2019), and are rated closely to real human speaker (Cambre et al., 2020). Despite these promising results, some dose of wary has to be adopted.

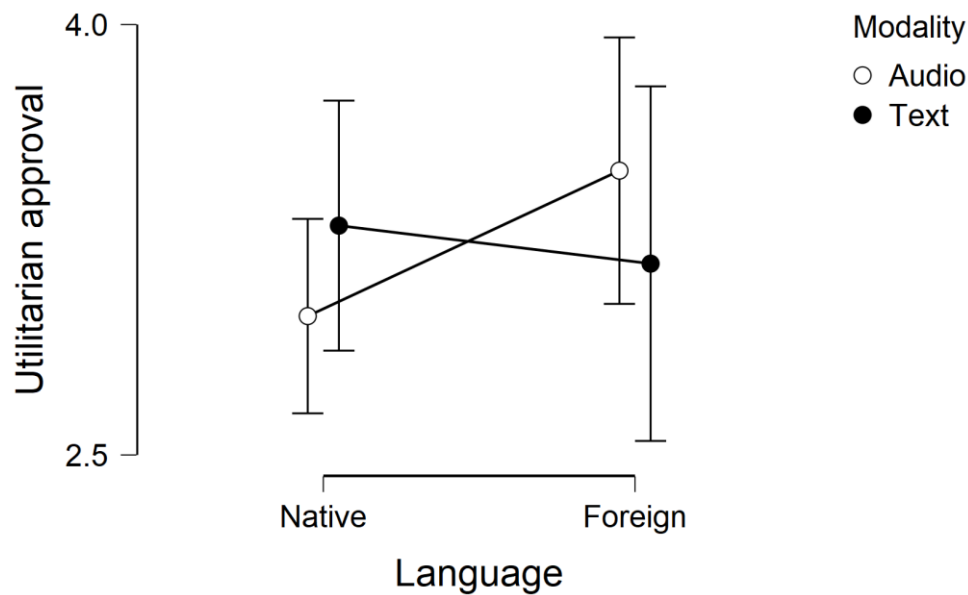
## Results

The analysis plan was preregistered. Before the main analysis, we averaged ratings of moral permissibility of action from three moral dilemmas and submitted them to a 2 (language: native vs. foreign) x 2 (modality: text vs. audio) between-subject ANOVA (Figure 2). Although visually similar to the ones observed in Brouwer (2019), the results revealed that neither the main effect of language, ( $F(1, 161) = 0.69, p = 0.408, \eta_p^2 = .004, B_{01} = 4.62$ ) nor the main effect of modality, ( $F(1, 161) < 0.001, p = 0.984, \eta_p^2 < 0.001, B_{01} = 7.36$ ) nor language by modality interaction, ( $F(1, 161) = 2.01, p = 0.159, \eta_p^2 = 0.012, B_{01} = 12.40$ ) were significant. In Bayesian terms, the evidence against main effects was moderate, while the evidence against the interaction was strong. Thus, our results suggest that use of a foreign language in the domain of moral judgments operates in the same way regardless of processing mode (i.e. reading or listening).

As preregistered, because we did not observe a significant effect of language when analyzing aggregated ratings, we then tested for differences for each item separately. For the footbridge dilemma, no effects were significant, all  $F < 1, B_{01} > 5.77$ . For the terrorist dilemma, we found no main effects,  $F < 1, B_{01} > 4$ , but a just significant language by modality interaction,  $F(1, 161) = 4.13, p = .044, \eta_p^2 = 0.025, B_{01} = 6.79$ . The interaction was, however, discounted in the Bayesian terms which returned moderate evidence against its existence. Finally, as in the footbridge dilemma, no effects were observed in the lost wallet dilemma: for language  $F(1, 161) = 1.54, p = .218, \eta_p^2 = 0.009, B_{01} = 3.67$ ; for modality,  $F(1, 161) = 1.15, p = .286, \eta_p^2 = 0.007, B_{01} = 4.38$ , and  $F < 1, B_{01} = 15.92$  or for the language by modality interaction. Hence, we found moderate to strong evidence against the language by modality interaction.

Figure 1

Utilitarian choices depending on language and mode of presentation in Experiment 1.



*Note.* Utilitarian approval is averaged permissibility score for three moral dilemmas. Error bars represent the 95% confidence interval. Dots represent data points.

## Discussion

After reanalyzing the data from a previous study (Brouwer, 2019) and collecting the data in an additional preregistered, high-powered experiment, we found that listening modality does not enhance the foreign language effect and it results in the same moral judgments as in the case of the text modality. However, although eyeballing the descriptive statistics could suggest its existence, neither of the datasets supported the existence of the foreign language effect. We therefore conclude that the moral foreign language effect may not be as robust as previously assumed. Most importantly, no differences between audio and text stimuli were observed. We therefore find no support for the amplification hypothesis (at least in cases where moral foreign language effect does not appear in the text modality), whereby audio processing of moral dilemmas increases the moral foreign language effect.

There are caveats in the interpretation of our findings. Despite the strong evidence against the modality effect, our results also do not support the foreign language effect. So, it is possible that in cases when foreign language effect is found, it is stronger in the audio modality. In contrast to the study by Brouwer (2019), we asked about moral appropriateness rather than about willingness to act. There is no consensus whether particular framing of the moral question does (Gawronski et al., 2017; Tassy et al., 2013) or does not differently affect the judgments much (O'Hara et al., 2010), and

some wariness is required. While modality by language interactions were non-significant, both the original experiment by Brouwer and our follow-up, showed stronger effects in the audio modality. Indeed, an amplification effect might exist, but may not be observed in samples of  $n < 165$ . The critical point for us is that even if the foreign language effect interacts with presentation modality, the interaction is weak so that it is unlikely to significantly alter the results of experiments.

Our findings have methodological and theoretical implications. From a methodological perspective, we provide evidence that future studies could implement auditory stimuli without major concerns of further affecting the moral foreign language effect. While this may appear a straightforward finding, it allows the foreign language effect research to go beyond the text-based laboratory settings. Our findings provide the basis for studies that mimic the real-world (better than text) and in which it might be easier to collect data with an auditory procedure, such as in field experiments, Virtual Reality simulations, (e.g. Francis et al., 2016, 2017), eye-tracking experiments (Godfroid, 2019) or visual world studies (Ito, Pickering, & Corley, 2018). Also, as speech is one of the most popular forms of communication, the potential presence of the foreign language effect in the auditory modality more broadly demonstrates how use of a foreign language might affect everyday decisions. For example: honest admitting to errors (by increased honesty, Bereby-Meyer et al., 2018); blaming an innocent person to protect a larger group (by permitting harm to be caused to others if more persons would benefit, Białek et al. 2019b; Muda et al., 2018); or higher willingness to invest in innovative and risky projects when analyzing and discussing the market data (by decreased risk aversion. For a broader view where the foreign language effect might be applicable, see Hadjichristidis, Geipel, & Surian, 2017).

A second implication of our experiment is theoretical. The moral foreign language effect was, again, not observed in the present study (as e.g., in Białek et al. 2019b; Muda et al., 2018; Brouwer 2019; Hayakawa et al., 2017). In research that has evidenced the effect, it was usually limited to the footbridge dilemma, with no strong evidence for it generalizing to other high-conflict dilemmas (Costa et al., 2014; Geipel et al., 2015; Nakamura, 2015). Hence, there appear to be unique features in some dilemmas that consistently cue more utilitarian responding, and this is critical in understanding how the moral foreign language effect and moral judgments in general work. The former seems rather obvious – by establishing what makes people change their choices, we can better explain the core mechanism of using a foreign language on decision making. This would have a broader application: once a feature shifting our morals is identified, we can better understand how moral judgments operate in general, and what element of our cognition drives utilitarian moral judgments. The latter seems to be one of the main goals of moral psychology (Białek, et al., 2019c).

Moreover, our results speak to another theoretical issue. Although speaking is more common than reading, and produces stronger affective reactions (Caldwell-Harris, 2014; Jankowiak & Korpai, 2018), the impact of using a foreign language on decision making seems to be unaffected by it. If the foreign language effect would be predominantly associated with affective processing, we would expect auditory presentation of the problem (speaking it) to amplify it (Campbell, 2005). This in turn should create more space for the foreign language effect to operate, similarly to greater affective responses to personal vs. impersonal moral dilemmas (Greene et al., 2001). And this should be even more pronounced in cases where the text-based foreign language effect does not work, as the room for a decision-switch should be greater. Since we observed no such effect, the affective explanation seems less plausible than the alternative ones such as reduced importance of consequences (Białek et al., 2019b) and reduced concern for making harm (Muda et al., 2018). This seems important, as in almost all studies regarding the foreign language effect, one can find suggestions that the foreign language effect works via attenuation of emotional reactions (e.g. Corey et al., 2017; Costa et al., 2014a, 2014b; Geipel, Hadjichristidis, & Surian, 2015a, 2015b, Hadjichristidis, Geipel, & Savadori, 2015; Hayakawa & Keysar, 2018). However, only a few studies have made an attempt to confirm that reduced emotionality is indeed the mechanism behind the impact of foreign language use on decisions and judgments. Results of these studies are mixed: two studies observed that reduced emotionality mediates the effect (Gao, Zika, Rogers, & Thierry, 2015; Hadjichristidis, Geipel, & Savadori, 2015), whereas the remaining studies revealed a lack of connection between emotions and the foreign language effect (Chan, Gu, Ng, & Tse, 2016; Geipel, Hadjichristidis, & Surian, 2015a, 2015b; Muda et al., 2020). Moreover, the issue is more complex because the effects of foreign language on low arousing problems are either null (Mækelæ, & Pfuhl, 2019; Muda et al., 2020; Vives, Aparici, & Costa, 2018) or negative (Białek et al., 2019a).

One of the limitations of our study is that we tested the foreign language effect under spoken and text stimuli in only one domain of the foreign language effect (i.e. the moral domain), whereas the foreign language effect impacts a variety of domains (e.g. risk-taking, risk perception, or illusory correlations (Keysar, Hayakawa, & An, 2012; Gao, Zika, Rogers, & Thierry, 2015; Hadjichristidis, Geipel, & Savadori, 2015). It may be the case that foreign language is affected differently by the modality of presentation in other domains of decision making. Another limitation of our conclusions is that we conducted only one experiment that differs from the Brouwer (2019) in several ways, i.e., we were asking for permissibility (vs. willingness to act) and using continuous (vs. dichotomous) scale. Note however, that our study was adequately powered, and re-analysis of Brouwer's data lead to the same conclusions. Another limitation connects with used procedure, i.e. with use of text-to-speech application (Ammazon Polly) that could be perceived as somewhat different than normal

human voice (Cambre et al., 2020) and thus, differently impact moral judgments. As we mentioned earlier, we decided for this move to avoid problems with accented speech (e.g. Crowther, Trofimovich, & Isaacs, 2016) but this could create other artefacts. Further work is needed to verify the modality effects in foreign language in real vs computer-generated speech.

Overall, our experiment, and the reanalysis of the original dataset from Brouwer (2019), found little to no differentiating effects of presenting the moral stimuli in auditory modality. Moreover, we haven't found evidence for the foreign language effect. As this is not the first research reporting null effects of the foreign language on moral judgments, further research is required to explain these null effects. Modality of presentation, however, is an unlikely candidate.

### Funding

The current project was financed by the resources of Polish National Science Centre (NCN) under Grant PRELUDIUM 2018/29/N/HS6/02058 assigned to RM. Work done by MB was supported by the National Science Centre, Poland (NCN) under Grant SONATA 2017/26/D/HS6/01159. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author contributions

All authors conceived the study. RM and DP designed the study protocol. DP gathered the data. RM and MB carried out the statistical analyses and drafted the manuscript. KF and DP provided critical revisions.

### References

- Białek, M., & De Neys, W. (2017). Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity. *Judgment and Decision Making*, 12(2), 148-167.
- Białek, 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
- Białek, M., Muda, R., Stewart, K. M., Niszczoła, P., & Pieńkosz, D. (2019a). Thinking in a foreign language distorts allocation of cognitive effort: Evidence from reasoning. *PsyArXiv*. DOI: 10.31219/osf.io/dnkc
- Białek, M., Paruzel-Czachura, M., & Gawronski, B. (2019b). 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
- Białek, M., Turpin, M. H., & Fugelsang, J. A. (2019c). What is the right question for moral psychology to answer? Commentary on Bostyn, Sevenhant, and Roets (2018). *Psychological science*, 30(9), 1383-1385. DOI: 10.1177/0956797618815171

- Bereby-Meyer, Y., Hayakawa, S., Shalvi, S., Corey, J. D., Costa, A., & Keysar, B. (2018). Honesty speaks a second language. *Topics in cognitive science*, 1-12. DOI: 10.1111/tops.12360
- Brouwer, S. (2019). The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *Journal of Multilingual and Multicultural Development*, 1-14. DOI: 10.1080/01434632.2019.1585863
- Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, 5, 1055. DOI: 10.3389/fpsyg.2014.01055
- Caldwell-Harris, C. L. (2015). Emotionality Differences Between a Native and Foreign Language Implications for Everyday Life. *Current Directions in Psychological Science*, 24, 214–219. DOI:10.1177/0963721414566268.
- Cambre, J., Maddock, J., Tsai, J., Colnago, J., & Kaye, J. Choice of Voices: A Large-Scale Evaluation of Text-to-Speech Voice Quality for Long-Form Content. Retrieved from: [https://juliacambre.com/TTS\\_Voice\\_Quality\\_CHI2020.pdf](https://juliacambre.com/TTS_Voice_Quality_CHI2020.pdf) DOI: 10.1145/3313831.3376789.
- Campbell, N. (2005). Getting to the heart of the matter: Speech as the expression of affect; rather than just text or language. *Language Resources and Evaluation*, 39(1), 109-118. DOI: 10.1007/s10579-005-2699-y
- Chan, Y.-L., Gu, X., Ng, J. C.-K., & Tse, C.-S. (2016). Effects of dilemma type, language, and emotion arousal on utilitarian vs. deontological choice to moral dilemmas in Chinese-English bilinguals. *Asian Journal of Social Psychology*, 19, 55–65. DOI: 10.1111/ajsp.12123
- Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43(7), 1109–1128. 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, 236–254. DOI:10.1016/j.cognition.2013.11.010.
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PloS ONE*, 9(4), e94842. DOI: 10.1371/journal.pone.0094842
- Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation*, 2(2), 160-182. DOI: 10.1075/jslp.2.2.02cro
- Dienes, Z. (2014). Using bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. DOI: 10.3389/fpsyg.2014.00781
- Everett, J. A., & Kahane, G. (2020). Switching Tracks? Towards a Multidimensional Model of Utilitarian Psychology. *Trends in Cognitive Sciences*, 24(2), 124-134. DOI: 10.1016/j.tics.2019.11.012
- Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39,175–191. DOI: 10.3758/BF03193146
- Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: transitioning from moral judgment to moral action?. *PloS ONE*, 11(10), e0164374. DOI: 10.1371/journal.pone.0164374

- Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific reports*, 7(1), 13954. DOI: 10.1038/s41598-017-13909-9.
- Gao, S., Zika, O., Rogers, R. D., & Thierry, G. (2015). Second Language Feedback Abolishes the “Hot Hand” Effect during Even-Probability Gambling. *The Journal of Neuroscience*, 35, 5983–5989. DOI: 10.1523/JNEUROSCI.3622-14.2015
- Gawinkowska, M., Paradowski, M. B., & Bilewicz, M. (2013). Second language as an exemptor from sociocultural norms. Emotion-related language choice revisited. *PLoS ONE*, 8,e81225. DOI:10.1371/journal.pone.0081225.
- Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology*, 113(3), 343-376. DOI: 10.1037/pspa0000086
- 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
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *American Statistician*, 60, 328–331. DOI: 10.1198/000313006X152649
- Godfroid, A. (2019). *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. New York. Routledge.
- Goodhue, D., Lewis, W., & Thompson, R. (2007). Research note—Statistical power in analyzing interaction effects: Questioning the advantage of PLS with product indicators. *Information Systems Research*, 18(2), 211-227. DOI: 10.1287/isre.1070.0123
- Goodwin, G., Piazza, J., & Rozin, P. (2014). Moral Character Predominates in Person Perception and Evaluation. *Journal of Personality and Social Psychology*, 106(1), 148-168. DOI:10.1037/a0034726
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108. DOI: 10.1126/science.1062872
- Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied*, 21, 117–129. DOI:10.1037/xap0000044.
- Hadjichristidis, C., Geipel, J., & Surian, L. (2017) How foreign language affects decisions: Rethinking the brain drain model. *Journal of International Business Studies*, 48(5). pp. 645-651. DOI: 10.1057/s41267-016-0040-1
- Hadjichristidis, C., Geipel, J., & Surian, L. (2017). Breaking magic: Foreign language suppresses superstition. *The Quarterly Journal of Experimental Psychology*, 24, 1-36. DOI:10.1080/17470218.2017.1371780



- Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*, 20(11), 791-793. DOI: 10.1016/j.tics.2016.08.004
- Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition*, 173, 8–15. DOI: 10.1016/j.cognition.2017.12.010
- Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, 98, 1-11. DOI: 10.1016/j.jml.2017.09.002
- Jankowiak, K., & Korpai, P. (2018). On modality effects in bilingual emotional language processing: Evidence from galvanic skin response. *Journal of Psycholinguistic Research*, 47(3), 663-677. DOI: 10.1007/s10936-017-9552-5
- Jeong, Y., Lee, J., & Kang, Y. (2019). Exploring Effects of Conversational Fillers on User Perception of Conversational Agents. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-6. DOI: 10.1145/3290607.3312913
- Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). ‘Utilitarian’ judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, 134, 193-209. DOI: 10.1016/j.cognition.2014.10.005
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23, 661–668. DOI:10.1177/0956797611432178
- Kroll, B. M. (1981) Developmental relationships between speaking and writing. In: B. M. Kroll & R. J. Van (Eds.) *Exploring Speaking-Writing Relationships: Connections and Contrasts*. (pp. 32–54). Urbana, IL: National Council of Teachers of English
- Lee, M.D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093-1096. DOI:10.1016/j.jesp.2010.05.025
- Liberman, A. M. (1992). The relation of speech to reading and writing. *Advances in Psychology*, 94, 167-178. DOI: 10.1016/S0166-4115(08)62794-6
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., ... & Matzke, D. (2019). JASP: graphical statistical software for common statistical designs. *Journal of Statistical Software*, 88(2), 1-17. DOI: 10.18637/jss.v088.i02
- Maekelae, M. J., & Pfuhl, G. (2019). Deliberate reasoning is not affected by language. *PLoS ONE*, 14(1), e0211428-e0211428.
- Moore, A. B., Louis Lee, N. Y., Clark, B. A. M., & Conway, A. R. A. (2011). In defense of the personal/impersonal distinction in moral psychology research: Cross-cultural validation of the dual process model of moral judgment. *Judgment and Decision Making*, 6, 186–195.
- Muda, R., P. Niszczoła, M. Białek, & Conway, P. (2018). “Reading Dilemmas in a Foreign Language Reduces Both Deontological and Utilitarian Response Tendencies.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44: 321–326. DOI: 10.1037/xlm0000447
- Muda, R., Walker, A. C., Pieńkosz, D., Fugelsang, J. A., & Białek, M. (2020). Foreign Language does not Affect Gambling-Related Judgments. *Journal of Gambling Studies*, 1-20. DOI: 10.1007/s10899-020-09933-6

- Nakamura, K. (2015). Effect is sure, but explanation is unsure: Closer investigation of the foreign language effect with Japanese participants. *Proceedings of the CogSci 2015*: Paper 0295. 1697-1702.
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, *14*, 1105–1107. DOI: 10.1038/nn.2886
- O'Hara, R. E., Sinnott-Armstrong, W., & Sinnott-Armstrong, N. A. (2010). Wording effects in moral judgments. *Judgment and Decision Making*, *5*(7), 547-554.
- Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes*, *109*(2), 134-141. DOI: 10.1016/j.obhdp.2009.03.002
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*(2), 165-184. DOI: 10.1023/A:1019923923537
- Spears, D., Fernández-Linsenbarth, I., Okan, Y., Ruz, M., & González, F. (2018). Disfluent fonts lead to more utilitarian decisions in moral dilemmas. *Psicológica Journal*, *39*(1), 41-63. DOI: 10.2478/psicolj-2018-0003
- Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in psychology*, *4*, 250. DOI: 10.3389/fpsyg.2013.00250
- Valkenburg, P. M., & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. *Journal of Adolescent Health*, *48*(2), 121-127. DOI: 10.1016/j.jadohealth.2010.08.020
- Vives, M.-L., Aparici, M., Costa, A. (2018). The limits of the foreign language effect on decision-making: The case of the outcome bias and the representativeness heuristic. *PLoS ONE* *13*(9): e0203528. DOI: 10.1371/journal.pone.0203528
- Wickelmaier, F. (2015). On not testing the foreign-language effect: A comment on Costa, Foucart, Arnon, Aparici, and Apesteguia (2014). *arXiv preprint arXiv:1506.07727*.
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, 1-61. DOI: 10.1017/S0140525X17001972