International Journal of Bilingual Education and Bilingualism

ISSN: 1367-0050 (Print) 1747-7522 (Online) Journal homepage: www.tandfonline.com/journals/rbeb20

The foreign language effect on moral judgement: insights from the self–other moral bias

Lu Jiao, Xiaohan Wang, Kalinka Timmer & Cong Liu

To cite this article: Lu Jiao, Xiaohan Wang, Kalinka Timmer & Cong Liu (2025) The foreign language effect on moral judgement: insights from the self–other moral bias, International Journal of Bilingual Education and Bilingualism, 28:4, 495-506, DOI: 10.1080/13670050.2024.2445260

To link to this article: https://doi.org/10.1080/13670050.2024.2445260

INTERNATIONAL JOURNAL OF BILINGUAL EDUCATION AND BILINGUALISM
2025, VOL. 28, NO. 4, 495–506
https://doi.org/10.1080/13670050.2024.2445260

The foreign language effect on moral judgement: insights from the self–other moral bias

Lu Jiao[a,b], Xiaohan Wang[a,c], Kalinka Timmer[d] and Cong Liu[a,b]

[a]Department of Psychology, School of Education Science, Qingdao University, Qingdao, People's Republic of China; [b]Brain, Cognition, and Language Learning Laboratory, Qingdao University, Qingdao, People's Republic of China; [c]School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK; [d]Faculty of Psychology, University of Warsaw, Warsaw, Poland

ABSTRACT

The moral foreign-language effect (MFLE) suggests biases present when making moral decisions in the native language are not present in the foreign language. However, the literature using explicit dilemmas shows inconsistent findings. The present study investigates whether MFLE has its origin in the reduced emotion hypothesis. Instead of the typically employed explicit paradigms, we utilize an implicit paradigm, avoiding conscious processing. Chinese–English bilinguals completed an implicit association test (Experiment 1) and an evaluative priming task (Experiment 2) in their native (L1: Chinese) and second language (L2: English). Both experiments found consistent evidence that the self–other moral bias was only observed in the native language. Therefore, we propose that the MFLE has its origin during the automatic associative stage. It results from the reduced emotional reaction in a foreign compared to the native language.

ARTICLE HISTORY
Received 6 August 2023
Accepted 16 December 2024

KEYWORDS
Foreign language effect; moral judgement; self–other moral bias; bilinguals

1. Introduction

Would you sacrifice one life by pushing them onto the track to save another five lives when an out-of-control trolley approaches the five people? When faced with such moral dilemmas, the automatic response is to say it's immoral to sacrifice one person's life to save more people (deontological judgment). On the other hand, more deliberative reasoning makes it more acceptable to sacrifice one person for the greater good of humanity (utilitarian decision) (Bartels et al. 2015). Thus, small changes in formulating such dilemmas impact people's moral choices (Corey et al. 2017). Furthermore, some recent studies have found that bilinguals are more likely to make utilitarian choices when the moral dilemmas are presented in a foreign language than when presented in the native language, known as the moral foreign language effect (MFLE) (Cipolletti, McFarlane, and Weissglass 2016; Costa, Dunabeitia, and Keysar 2019; Del Maschio et al. 2022; Geipel, Hadjichristidis, and Surian 2015; for a review, see Purpuri et al. 2024a). However, not all studies showed an MFLE (for a review, see Stankovic, Biedermann, and Hamamura 2022). These studies mainly used various explicit dilemmas to examine the MFLE (Białek, Paruzel-Czachura, and Gawronski 2019; Brouwer 2019; Geipel, Hadjichristidis, and Surian 2015; Hayakawa et al. 2017). As both the type of dilemma and the language can impact moral judgements (i.e. making utilitarian choices more likely), it is unclear why the MFLE

CONTACT Cong Liu congu902@gmail.com

revealed inconsistent effects. Therefore, we aimed to examine whether the MFLE is present using implicit measures (i.e. implicit association test in Experiment 1 and evaluative priming task in Experiment 2). Implicit measures index automatic processing, which avoids deliberative reasoning.

1.1. Foreign language effect on moral decision-making

The initial investigation on MFLE was conducted by Costa et al. (2014b), in which they used the footbridge dilemma. In this dilemma, an oncoming trolley is about to kill five people tied to the track. The only way to save them is to push a bystander off the footbridge onto the tracks, killing him but stopping the trolley. Monolingual research shows that most people are not willing to sacrifice an innocent bystander to save five people (e.g. Greene et al. 2001). Costa et al. (2014b) replicated this in bilinguals, with only 18% willing to sacrifice the bystander when the dilemma was presented in their native language. Crucially, 44% of the bilinguals agreed to sacrifice the bystander when the dilemma was presented in their foreign language. In this type of dilemma, the MFLE was replicated with different bilingual populations (Brouwer 2021; Corey et al. 2017; Romero-Rivas, López-Benítez, and Rodríguez-Cuadrado 2020).

Now consider a similar problem, the switch dilemma. As before, a trolley threatens to kill five people. Now the only way to save the five people is by pulling a lever to switch the trolley onto an alternate track. Unfortunately, one man is lying on the alternate track, killing him. Now, most people are willing to sacrifice one person to save five (utilitarian decision) even though the action in the footbridge and switch dilemmas have the same outcome (e.g. Greene et al. 2001). The difference between the footbridge and switch dilemma is whether it involves exerting personal force from pushing the man or indirect consequences from switching a lever (Corey et al. 2017). One of the explanations is that people consider their personal involvement as more important in the footbridge compared to the switch dilemma (Lieberman et al. 2007). While the MFLE was present for more personal moral dilemmas (i.e. an increase in the choice to save a larger number of people), native/foreign language did not modulate impersonal moral dilemmas (Brouwer 2021; Costa et al. 2014b; Geipel, Hadjichristidis, and Surian 2015). To conclude, the MFLE is not shown in all studies and could depend on the specific context a dilemma is presented in (for a review, see Stankovic, Biedermann, and Hamamura 2022).

Other small contextual changes also modulated whether the MFLE was present or not. For example, Chan et al. (2016) presented participants with 22 personal and 17 impersonal dilemmas to investigate the presence/absence of the MFLE. They did not find an MFLE in any dilemma other than the personal Footbridge dilemma. These findings indicated that the foreign language effect on moral decisions has potential boundaries (Čavar and Tytus 2018; Del Maschio et al. 2022; Dylman and Champoux-Larsson 2020). Nadarevic, Klein, and Dierolf (2021) also found inconsistent language effects using a battery of dilemmas. They revealed that people's moral judgements are driven by different parameters based on a model-based analysis: people's sensitivity to consequences (C-parameter), their sensitivity to norms (N-parameter) and their general preference for action or inaction (I-parameter). Thus, this analysis suggests that the inconsistent FMLE results might be related to a focus on different parameters during the deliberative reasoning of different dilemma sets used in each study. This makes the FMLE effect fragile and context-specific (Muda et al. 2017; Nadarevic, Klein, and Dierolf 2021).

The studies mentioned above have two major limitations (Barabadi, Rahmani Tabar, and Booth 2021; Costa, Vives, and Corey 2017). First, most former studies on MFLE mainly used the traditional dilemma approach, which heavily depended on specific dilemmas that are context-dependent (Brouwer 2019, 2021; Cipolletti, McFarlane, and Weissglass 2016; Costa et al. 2014b; Driver 2020; Dylman and Champoux-Larsson 2020; Geipel, Hadjichristidis, and Surian 2015). The framing of moral questions has modulated the presence/absence of the MFLE (Del Maschio et al. 2022). Other replicability issues could be due to yet unknown contextual differences. Second, the explicit nature of the traditional dilemma approach makes participants were directly asked about their

choices in different moral dilemmas. This self-report might lack objectivity because individuals can fake their responses.

The present study addresses the above limitations by employing implicit measures to investigate the MFLE at a purely conceptual level, independent of the specific context of dilemmas. Specifically, we focused on the self–other moral bias in two classical implicit tasks: Implicit Association Test (IAT) and Evaluative Priming Task (EPT). The self–other moral bias showed that people hold a stronger association between *self* and *deontology* (or *other* and *utilitarian*) than *other* and *deontology* (*self* and *utilitarian*). This bias has been observed in the implicit tasks (i.e. IAT and EPT), which are less likely to be influenced by context than explicit tasks (Li, Li, and Rao 2021). In addition, individuals are less likely to fake their responses to implicit tasks.

1.2. Mechanisms underlying the MFLE

In previous studies, moral decisions were always made by asking participants about their choices in different moral dilemmas. The Dual-Process theory assumes that two distinct systems determine how we make moral decision making: one involving an emotional process that is relatively quick, intuitive, and automatic (i.e. System 1) and another involving a rational process that is relatively slow, deliberate, and controlled (i.e. System 2) (Greene et al. 2001, 2008; Kahneman 2003). Accordingly, two hypotheses (i.e. the *reduced emotion hypothesis* and the *increased deliberation hypothesis*) have been proposed, indicating two different mechanisms potentially underlying the MFLE (Conway and Gawronski 2013; Del Maschio et al. 2022). While the *reduced emotion hypothesis* assumes that the FMLE results from the reduction of emotion resonance associated with System 1 processing, the *increased deliberation hypothesis* assumes that this effect results from increased deliberative thinking associated with System 2 processing. The traditional dilemma approach cannot separate between the *reduced emotion hypothesis* and the *increased deliberation hypothesis*; Cipolletti, McFarlane, and Weissglass 2016; Costa et al. 2014b. A recent study using the Process Dissociation technique, which was developed to separate different processes (i.e. emotional and deliberate processes) involved in moral judgement, suggests that the MFLE results from emotional reactions instead of increased deliberation (Hayakawa et al. 2017). However, their findings were based on specific dilemmas and, thus, still context-dependent. Therefore, it's still unclear whether the MFLE primarily reflects reduced emotional reactions based on System 1 or increased deliberative thinking based on System 2. The present study investigates whether MFLE has its origin in the reduced emotion hypothesis with implicit paradigms, which avoid explicit reasoning.

1.3. The present study

The present study investigated whether using a foreign language, as opposed to the native language, affects the self–other bias in moral judgments. We asked whether the MFLE exists at the purely conceptual level without top-down processes and is thus independent of specific contexts. Due to the ongoing replication crisis in psychology research (Maxwell, Lau, and Howard 2015), two classical implicit tasks (i.e. IAT and EPT) were employed to verify the robustness of our findings in unbalanced Chinese–English bilinguals. Both implicit tasks in the two experiments were separate in the native language (i.e. Chinese) and the foreign language (i.e. English). In Experiment 1 (IAT), unbalanced bilinguals assigned each stimulus word to one of two or four categories (*Self* or *Other* and *Deontological* or *Utilitarian*). In Experiment 2, another group of unbalanced bilinguals judged whether target words were *Deontological* or *Utilitarian*. Target words were preceded by prime words from the *Self* or *Other* category, which had a congruent (*Self–Deontological*) or incongruent relationship to the target (*Other–Utilitarian*).

If foreign language influences moral judgment during the automatic associative stage (System 1), the self–other moral bias will differ across two language contexts. Based on the findings of Li, Li, and Rao (2021), we expected the self–other moral bias would be present in the Chinese (native language)

condition but reduced or absent in the English (foreign language) condition. As implicit measures are suggested to only employ automatic associative processes (i.e. System 1) (Gawronski, De Houwer, and Sherman 2020), we speculate that the observed MFLE was driven by reduced emotional reaction instead of increased deliberation.

Crucially, while both tasks are widely used paradigms for studying implicit cognition, their unique methodologies tap into different mechanisms. IAT measures congruence in a categorization task in which participants must simultaneously process and sort two types of categories. The IAT's reliance on explicit categorization makes it highly reliable, with robust findings across studies. However, this explicit nature may allow some conscious control due to explicit cognitive evaluation (Yamaguchi and Beattie 2020). EPT only requires categorization of the target word, while the primes are implicitly related to the target and, in turn, categories. Therefore, EPT's indirect approach focuses more on automatic processing through implicit associations. However, this also makes it more prone to noise and variability due to the dependence on implicit effects. Therefore, to verify the robustness of the findings in Experiment 1 and exclude the possible explicit process in IAT, another group of unbalanced Chinese-English bilinguals completed an EPT task, which depends to a greater extent on implicit processing.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty participants (10 males) from Qingdao University participated in experiment 1, which the ethical committee of the same University had approved. The mean age of the participants was 19.90 ± 1.35. They all signed a written informed consent form before their participation and received payment after participation. All participants were born in China and had no experience studying abroad. Chinese is the participants' native language. Their average age of starting to learn English is 8.14 (SD = 2.51). They rated their proficiency in both languages for listening, speaking, reading, and writing on a seven-point Likert scale of 1–7, with 1 indicating the lowest and 7 indicating the highest (Liu et al. 2021). Paired-samples *t*-tests showed that the proficiency ratings for all language skills were significantly higher in Chinese than in English (all $ts > 7.131$, all $ps < 0.001$, see Table 1).

2.1.2. Materials

In IAT, we presented the same 20 Chinese words and 20 English words. The words were divided into four words related to oneself, four related to others, six deontological words, and six utilitarian words (see Table 2). The words are good representatives of the four categories (Li, Li, and Rao 2021). Twenty raters who did not know the purpose of the experiment rated their familiarity with these words on a seven-point Likert scale (7 indicates most familiar and 1 indicates least familiar). Independent-samples *t*-tests indicated no significant differences in familiarity between the 20 Chinese and 20 English words ($t = 0.098$, $p = 0.922$).

Table 1. Means (and SDs) for self-reported language proficiency ratings and age of acquisition (AoA) for both Chinese and English in both experiments.

	Experiment 1		Experiment 2	
Self-ratings	Chinese	English	Chinese	English
AoA		8.43 (1.98)		8.14 (2.51)
Listening	6.33 (0.96)	2.93 (1.34)	6.14 (1.15)	3.36 (1.34)
Speaking	5.67 (1.52)	2.93 (1.23)	5.86 (1.27)	3.29 (1.18)
Reading	6.07 (1.26)	4.00 (1.37)	6.00 (1.16)	4.29 (1.46)
Writing	5.80 (1.35)	3.60 (1.19)	5.64 (1.22)	3.43 (1.14)

Table 2. Stimulus words presented in both Experiment 1 and Experiment 2.

	Category	Stimuli words
Chinese words	Self	我, 自我, 我自己, 我本人
	Other	他, 别人, 其他人, 陌生人
	Deontological	正义, 道德, 品德, 公德, 公平, 良心
	Utilitarian	效益, 利益, 收益, 好处, 回报, 效率
English words	Self	I, self, myself, my
	Other	he, other, she, stranger
	Deontological	equity, fairness, honest, justice, moral, principled
	Utilitarian	advantage, benefit, effective, gain, productive, profit

2.1.3. Procedure

The IAT is a classical task for measuring the strengths of automatic associations between words and categories. Before IAT was started, the experimenter explained the concept of self/other and deontology/utilitarianism to the participants. Then, the participants familiarized themselves with all the words. During the experiment, participants assigned each stimulus word to one of two categories (in blocks 1, 2, and 5) or one of four categories (in blocks 3, 4, 6, and 7) by pressing E (left hand) or I (right hand)on the keyboard as quickly as possible.

The categories were presented on the top-left and top-right of the screen and remained throughout each block. Each stimulus word was presented in the center of the screen until a categorization decision was made. When participants chose an incorrect category, a red cross appeared on the screen until the correct response was given. The next trial began after the presentation of a blank screen for 250 ms.

As shown in Figure 1, the IAT procedure consisted of seven blocks. In Block 1 (20 trials), participants categorized the self/other words into the Self and Other categories. Block 2 (20 trials) required participants to categorize deontological/utilitarian words into the Deontological and Utilitarian categories. In Blocks 3 (20 trials) and 4 (40 trials), both the 'Self–Other' and 'Deontological–Utilitarian' categories were presented (two on each side of the screen), in which participants categorized each word into its corresponding category (e.g. Self). Block 5 was identical to Block 1 but with labels in the reverse positions of Block 1. Blocks 6 and 7 were identical to Blocks 3–4, except that the positions of the 'Self–Other' categories were reversed (see Figure 1). This procedure in the present study consisted of two parts with the same design, one in Chinese and one in English.

2.1.4. Design

The factors of Language (Chinese vs. English) and Congruency (congruent vs. incongruent) were manipulated using a within-subjects design in the experiment. In the combined blocks (i.e. blocks 3, 4, 6, and 7), we had the same number of congruent and in congruent pairings. Congruent pairings are *Self + Deontological* and *Other + Utilitarian*, and incongruent pairings are *Self + Utilitarian* and *Other + Deontological*. The blocks of incongruent and congruent pairings were counterbalanced across participants. The order of the languages was also counterbalanced across participants. The word stimuli were presented in random order.

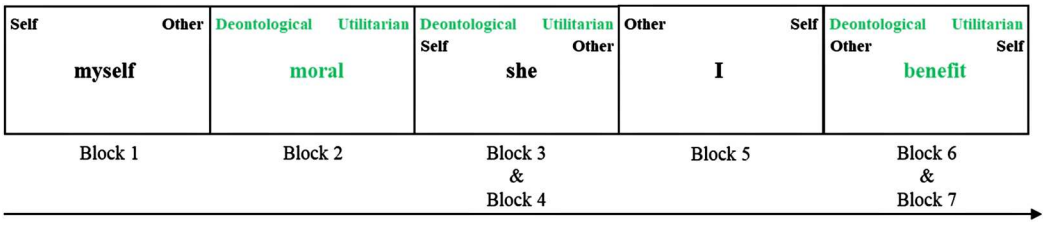

Figure 1. Schematic procedure of the implicit association test for the English part. Only self/other words were presented in blocks 1 and 5, and only deontological/utilitarian words were presented in block 2. By contrast, in combined blocks of 3, 4, 6, and 7, a self/other word or a deontological/utilitarian word was presented as a target word in random order.

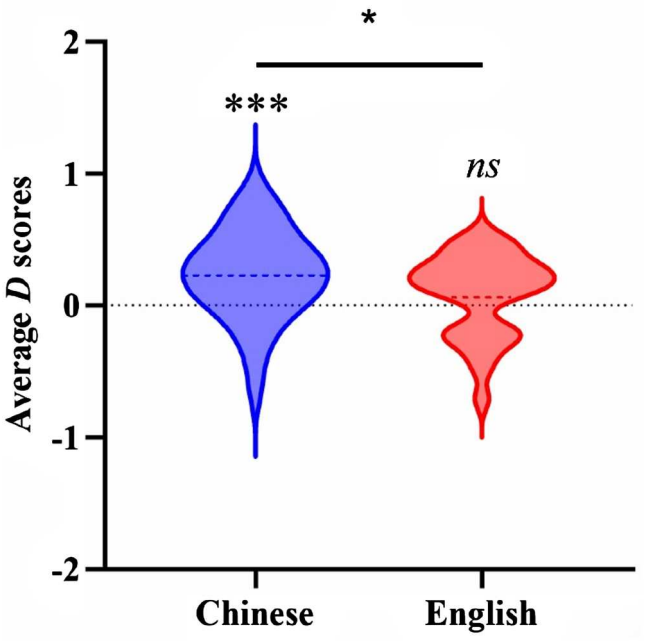

Figure 2. *D* scores in each Language (Chinese vs. English) context. The colored dotted lines denote means in the Chinese and English context. Note: *ns* indicates non-significant; * $p < .05$; *** $p < .001$.

2.2. Result and discussion

Response latencies from combined blocks (Blocks 3, 4, 6, and 7) were recorded and analyzed using the improved IAT scoring algorithm (Greenwald, Nosek, and Banaji 2003). *D* scores were calculated by dividing the differences in the mean scores between the congruent and incongruent IAT blocks by the standard deviation of all the trials in all blocks. A positive *D* score represents a faster response to congruent pairings (i.e. *Self* + *Deontological* and *Other* + *Utilitarian*) than to incongruent pairings (i.e. *Self* + *Utilitarian* and *Other* + *Deontological*), indicating a self–other moral bias.

In the Chinese context, the *D* score was significantly higher than zero, $D = 0.23$, $t\ (29) = 3.872$, $p = 0.001$, 95% confidence interval [0.11, 0.36], suggesting that the participants were inclined to react more rapidly to congruent pairings than to incongruent pairings. This result demonstrated that deontological words were more strongly associated with words related to the self than words related to others and vice versa for utilitarian words (i.e. self–other moral bias). In contrast, in the English condition, the difference between the *D* score and zero was not significant, $D = 0.09$, $t\ (29) = 1.555$, $p = 0.131$, 95% confidence interval [−0.03, 0.20], suggesting the response latencies in congruent pairings were similar to incongruent pairings. This result demonstrated that the self–other moral bias was absent in the English condition. A paired-sampled *t*-test supported this finding with a significantly larger *D* score in Chinese than in English, $t\ (29) = 2.202$, $p = 0.036$, 95% confidence interval [0.01, 0.28] (see Figure 2). These findings suggested that the self–other moral bias was influenced by the language in which the task was performed (native vs. foreign).

3. Experiment 2

In Experiment 1, the IAT results showed that the self–other bias existed in Chinese but not in English. Considering that the manifestation of implicit biases in the IAT depends on explicit categorization involving some degree of conscious control (Yamaguchi and Beattie 2020), we employed another widely used implicit paradigm (i.e. EPT) in Experiment 2 (Gawronski, De Houwer, and Sherman 2020). Experiment 2 aimed to replicate and verify the findings of Experiment 1.

3.1. Method

3.1.1. Participants

Thirty-six participants (11 males) from Qingdao University were paid to participate in Experiment 2, which the ethical committee of the same University approved. None had participated in Experiment 1. The mean age of the participants was 19.29 ± 1.08. All participants signed a written informed consent form and received payment for their participation. They also completed the same self-rating questionnaire as in Experiment 1. Paired-samples *t*-tests showed that the proficiency ratings for all language skills (i.e. listening, speaking, reading, and writing) were significantly higher in L1 than in L2 (all $ts > 4.869$, all $ps < 0.001$, see Table 1).

3.1.2. Materials

The materials were the same as in Experiment 1 (see Table 2).

3.1.3. Procedure

The EPT is based on the principle of evaluative priming, where exposure to a prime influences the speed and accuracy of evaluating a subsequent target stimulus. The experimenter explained the concepts of self, other, deontology, and utilitarianism to the participants. The participants familiarized themselves with the prime words (four words related to the concepts of *self* and *others*) and target words (six deontological and six utilitarian words) to help them clarify the correct classification of the words. During the experiment, participants were asked to read the prime word aloud to ensure that they paid attention to the prime words (Fazio et al. 1986; Li, Li, and Rao 2021). Next, they were instructed to decide whether the target word was *deontological* or *utilitarian* by pressing E or I on the keyboard as quickly as possible. Each trial began with a red fixation cross presented for 500 ms, followed by a prime word for 200 ms. Then, a 100 ms blank screen appeared, followed by the presentation of a target word until a response was given (see Figure 3). The EPT consisted of two blocks with the same design, one in Chinese and one in English. Twelve practice trials helped participants understand the task before starting the experiment.

3.1.4. Design

As in Experiment 1, the factors of Language (Chinese vs. English) and Congruency (congruent vs. incongruent) were manipulated with a within-subjects design in Experiment 2. The order of Language blocks was counterbalanced between participants (96 trials in each language context). Within each Language context, half of the trials were congruent, and the other half were incongruent (48 trials each). In the congruent condition, self-primes were combined with deontological targets and other-primes with utilitarian targets. In the incongruent condition, the opposite combinations were presented (e.g. self with utilitarian). Each prime word was paired once with a target word from the two categories and presented randomly in each block. Lastly, the position of the response keys (left and right) were counterbalanced across participants.

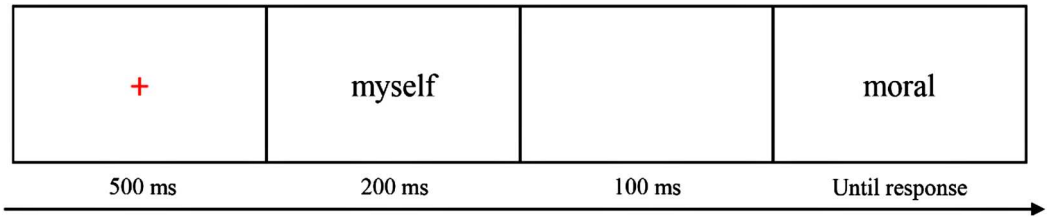

Figure 3. An example of a trial within the evaluative priming task.

3.2. Result and discussion

Response latencies to the targets were recorded and analyzed to examine the self–other moral bias. In Experiment 2, all participants had an accuracy above 80%. Trials with incorrect responses were excluded from the analyses. Absolute outliers (trials with response latencies beyond 250–2500 ms) and relative outliers (trials with a response latency over 2SD from the mean per condition) were also removed from the analyses. This left 87% of the data in the analyses.

The 2 (Chinese vs. English) by 2 (congruent vs. incongruent) repeated-measures ANOVA revealed that the main effect of Language was significant, $F(1, 34) = 17.454$, $p < 0.001$, $\eta_p^2 = 0.333$, with faster response in the Chinese context (M = 778 ms) than the English context (M = 867 ms, $p < 0.001$). The main effect of Condition was not significant, $F(1, 34) = 3.028$, $p = 0.082$, $\eta_p^2 = 0.084$. Critically, the interaction between Language and Condition reached significance, $F(1, 34) = 6.085$, $p = 0.019$, $\eta_p^2 = 0.148$. Post hoc analyses revealed that the latencies of the congruent condition (M = 763 ms) were significantly shorter than that of the incongruent condition (M = 786 ms, $p < 0.001$) in the Chinese context. In contrast, in the English context, the latencies in the congruent condition (M = 868 ms) were similar to those in the incongruent condition (M = 867 ms, $p = 0.907$) (see Figure 4).

These findings suggest a self–other moral bias was only observed in the Chinese context. In other words, participants associated the concept of self with deontology to a greater extent than the concept of other. Furthermore, the concept of other was more related to utilitarianism than the concept of self. This self–other moral bias was absent in English. Overall, these findings confirmed the results of Experiment 1, showing language context influenced the self–other moral bias.

4. General discussion

The present study investigated whether the foreign language effect modulates moral judgement during which people judge implicit associations between *self and deontology* and between *other and utilitarianism*. We introduced two implicit tasks (i.e. the IAT and the EPT) to investigate if the

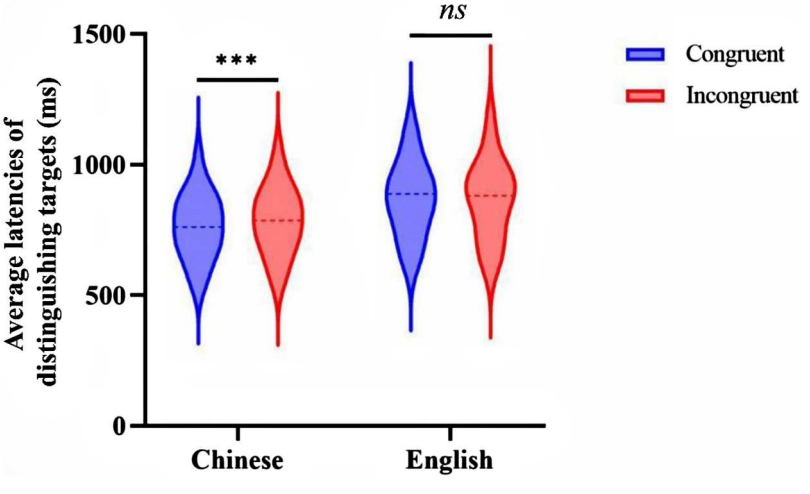

Figure 4. The average response latencies to categorize target words in the congruent and incongruent condition in each Language (Chinese vs. English). The colored dotted lines denote means in different conditions. Note: *ns* indicates non-significant; *** $p < .001$.

self–other moral bias was influenced by the language in which words were presented (native vs. foreign). In two experiments, we found consistent evidence that the self–other moral bias was present in the native language but absent in the foreign language. This suggests that in the native language context, participants associated utilitarian words to a greater degree with the concept of other than self. And vice versa, deontological words were more related to the concept of self than others. Thus, during automatic associative tasks, the foreign language influences moral judgment compared to the native language.

Although previous research has demonstrated that the foreign language can affect moral judgments in some specific moral dilemmas (Brouwer 2021; Hayakawa et al. 2017; Romero-Rivas, López-Benítez, and Rodríguez-Cuadrado 2020), there's no consensus on whether MFLE is a replicable effect (see a review, Stankovic, Biedermann, and Hamamura 2022). Researchers have indicated that the divergent results concerning MFLE may stem from different types of reasoning used during various moral scenarios/dilemmas (Barabadi, Booth, and Rahmani Tabar 2023). For example, Chan et al. (2016) employed 39 dilemmas to investigate the MFLE, but they only found an MFLE in the Footbridge dilemma. Others also found inconsistent language effects in a battery of dilemmas (Białek, Paruzel-Czachura, and Gawronski 2019; Hayakawa et al. 2017; Muda et al. 2017; Nadarevic, Klein, and Dierolf 2021). Model-based analysis revealed that depending on the explicit dilemmas different parameters (e.g. focusing on consequences, norms, or (in)action) are employed by participants during deliberative reasoning.

The crucial impact deliberate reasoning might have on moral decisions made us seek new ways to investigate the MFLE. While previous studies mainly used explicit measurement (i.e. traditional dilemma approach) to investigate the MFLE, the present study investigated the MFLE by introducing and applying classical implicit paradigms. Compared to the explicit measurement involving both automatic emotional process (i.e. System 1) and controlled rational process (i.e. System 2), the present study's implicit paradigms only focus on System 1. Furthermore, individuals are less likely to fake their responses during implicit tasks only, depending on the automatic emotional process. This is important as the dilemmas presented in explicit paradigms are often extreme and individuals are not likely to encounter them in their daily life. As the present study revealed the MFLE emerged in implicit tasks, we suggest that the origin of MFLE is during the automatic processing stage of System 1 and that different strategies applied during the deliberative reasoning stage of System 2 can distort MFLE effects. Therefore, we propose implicit tasks could, over time, prove more reliable than specific moral dilemmas. Overall, the findings in the present study provided direct evidence supporting the existence of MFLE from a new perspective.

Our findings with implicit paradigms supported the *reduced emotion hypothesis*, which proposes that the mechanism behind the foreign language effect is the reduction in emotional processing when using a foreign language (Costa et al. 2014b; Geipel, Hadjichristidis, and Surian 2015; Keysar, Hayakawa, and An 2012). We observed that the self–other moral bias in the native language was absent in the foreign language context. This is in line with a previous study using the Process Dissociation technique to investigate the mechanism underlying the MFLE (Hayakawa et al. 2017). They modulated various explicit moral dilemmas so that they could distinguish between deontological and utilitarian responses for each participant. Participants were found to reduce the number of their deontological responses when using a foreign language and did not show changes in utilitarian choices. As deontological judgements have been related to automatic emotional processing during System 1 (Bartels et al. 2015), Hayakawa et al. (2017) suggest MFLE is related to reduced emotional processing in the foreign language and not to increased deliberation. Thus, we argue that the MFLE was driven by reducing emotional reaction instead of increased deliberation. Specifically, we suggest that if emotional arousal is reduced in the foreign language context compared to the native language, participants make decisions with fewer cognitive biases.

The present finding that the self–other bias present in the native language is not in a foreign language could also be in line with broader findings in the foreign language effect literature. For example, biases such as the framing effect (i.e. people are biased to picking options they view as

a gain instead of loss, even if both lead to the same result; Holleman, Kamoen, and Struiksma 2021; Keysar, Hayakawa, and An 2012; Liu et al. 2022; Winskel et al. 2016), the loss aversion bias (i.e. people tend to take more risks when the same problem is framed in terms of losses than in terms of gains; Costa et al. 2014a), and the hot hand fallacy (i.e. the belief that a continuous series of positive results will continue to thrive; Gao et al. 2015) disappear in a foreign language. Especially when a foreign language is acquired through formal instruction instead of immersion, the emotional resonance in one's foreign language is weaker than in one's native language (Dewaele 2010). Consequently, emotional blunting in foreign languages is suggested to reduce cognitive biases. Thus, our findings on moral decisions might come from a more general reduction of emotional processing affecting different types of decisions.

Although the present study offers new insights into the relationship between the foreign language effect and moral judgments, several limitations should be addressed in future research. First, we did not control for the part of speech of the experimental words, which might influence the response latencies to some degree, especially when comparing the participants' response latencies between Chinese and English words. Second, previous studies have shown that men tend to be more utilitarian whereas women tend to be more sensitive and emotional when making moral judgments (Wei 2023). Having more balanced groups of male and female participants would be better. Third, participants only self-rated their L2 proficiency in the present study. It would be beneficial to use more objective measures of L2 proficiency to ensure greater accuracy and consistency in assessing the impact of L2 proficiency on moral judgement in future studies (Purpuri et al. 2024b).

5. Conclusion

In conclusion, the current study's findings revealed a foreign language effect on moral judgements during automatic bottom-up decision making that was based on System 1. This is evidenced by the alteration of the self–other moral bias in a foreign language context as opposed to the native language. This effect is most likely attributed to reduced emotional arousal in the foreign language, as implicit compared to explicit paradigms are less likely to require deliberative reasoning. Furthermore, in implicit measurement tasks, it is not so easy to fake your response as in explicit decisions (Li, Li, and Rao 2021). To the best of our knowledge, this is the first empirical study using implicit measurements to examine the MFLE. Overall, the present study extends the research on the MFLE from the contextual level (explicit dilemmas) to the conceptual level (subconscious associations), paving a novel avenue to understanding how a foreign language can affect moral judgments.

Disclosure statement

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

Funding

The study is supported by the National Natural Science Foundation of China (62107024), the Humanity and Social Science Youth foundation of Ministry of Education in China (22YJC190015), the Natural Science Foundation of Shandong Province in China (ZR2021QF012), the Young Innovative Research Team in Shandong Higher Education Institutions in China (2022RW030), and the OPUS 23 grant (2022/45/B/HS6/01931) from the National Science Center of Poland (NCN).

Notes on contributors

Lu Jiao is currently an associate professor at Qingdao University (China). Her research interests include foreign language effect, second language learning and co-learning.

Xiaohan Wang used to be an undergraduate student at Qingdao University (China). Now she pursues her Master's degree at the University of Edinburgh.

Kalinka Timmer is an associate professor at the University of Warsaw (Poland). Her research interests include bilingualism (bilingual control), cognitive control, and second language acquisition.

Cong Liu is an associate professor of Psychology at Qingdao University (China). His research interests include second language learning, bilingual language switching, and foreign language effects on social decisions.

References

Barabadi, E., Tabar M. R., and Booth J. R. 2023. "The effects of foreign language and religiosity on moral decisions: Manipulating norms and consequences ." *Journal of Cognition and Culture, 23* (3-4): 310–337. https://doi.org/10.1163/15685373-12340165.

Barabadi, E., M. Rahmani Tabar, and J. R. Booth. 2021. "The Relation of Language Context and Religiosity to Trilemma Judgments." *Journal of Cross-Cultural Psychology* 52 (6): 583–602. https://doi.org/10.1177/00220221211033987.

Bartels, D. M., C. W. Bauman, F. A. Cushman, D. A. Pizarro, and A. P. McGraw. 2015. "Moral Judgment and Decision Making." In *The Wiley Blackwell handbook of judgment and decision making*, edited by G. Keren and G. Wu, 478–515. Chichester, England: Wiley. https://doi.org/10.1002/9781118468333.ch17.

Białek, M., M. Paruzel-Czachura, and B. Gawronski. 2019. "Foreign Language Effects on Moral Dilemma Judgments: An Analysis Using the CNI Model." *Journal of Experimental Social Psychology* 85:103855. https://doi.org/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. https://doi.org/10.1080/01434632.2019.1585863.

Brouwer, S. 2021. "The Interplay Between Emotion and Modality in the Foreign-Language Effect on Moral Decision Making." *Bilingualism: Language and Cognition*, 24 (2): 223–230. https://doi.org/10.1017/s136672892000022x.

Čavar, F., and A. E. Tytus. 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. https://doi.org/10.1080/01434632.2017.1304397.

Chan, Y. L., X. Gu, J. C. K. Ng, and C. S. Tse. 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 (1): 55–65. https://doi.org/10.1111/ajsp.12123.

Cipolletti, H., S. McFarlane, and C. Weissglass. 2016. "The Moral Foreign-Language Effect." *Philosophical Psychology* 29 (1): 23–40. https://doi.org/10.1080/09515089.2014.993063.

Conway, P., and B. Gawronski. 2013. "Deontological and Utilitarian Inclinations in Moral Decision Making: A Process Dissociation Approach." *Journal of Personality and Social Psychology* 104 (2): 216–235. https://doi.org/10.1037/a0031021.

Corey, J. D., S. Hayakawa, A. Foucart, M. Aparici, J. Botella, A. Costa, and B. Keysar. 2017. "Our Moral Choices are Foreign to us." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 43 (7): 1109–1128. https://doi.org/10.1037/xlm0000356.

Costa, A., J. A. Dunabeitia, and B. Keysar. 2019. "Language Context and Decision-Making: Challenges and Advances." *Quarterly Journal of Experimental Psychology* 72 (1): 1–2. https://doi.org/10.1177/1747021818789799.

Costa, A., A. Foucart, I. Arnon, M. Aparici, and J. Apesteguia. 2014a. "Piensa" Twice: On the Foreign Language Effect in Decision Making." *Cognition* 130 (2): 236–254. https://doi.org/10.1016/j.cognition.2013.11.010.

Costa, A., A. Foucart, S. Hayakawa, M. Aparici, J. Apesteguia, J. Heafner, and B. Keysar. 2014b. "Your Morals Depend on Language." *PLoS One* 9 (4): e94842. https://doi.org/10.1371/journal.pone.0094842.

Costa, A., M. Vives, and J. D. Corey. 2017. "On Language Processing Shaping Decision Making." *Current Directions in Psychological Science* 26 (2): 146–151. https://doi.org/10.1177/0963721416680263.

Del Maschio, N., G. Del Mauro, C. Bellini, J. Abutalebi, and S. Sulpizio. 2022. "Foreign to Whom? Constraining the Moral Foreign Language Effect on Bilinguals' Language Experience." *Language and Cognition* 14 (4): 511–533. https://doi.org/10.1017/langcog.2022.14.

Dewaele, J.-M. 2010. "Multilingualism and Affordances: Variation in Self-Perceived Communicative Competence and Communicative Anxiety in French L1, L2, L3 and L4." *International Journal of Applied Linguistics* 48:105–129. https://doi.org/10.1515/iral.2010.006.

Driver, M. Y. 2020. "Switching Codes and Shifting Morals: How Code-Switching and Emotion Affect Moral Judgment." *International Journal of Bilingual Education and Bilingualism* 25 (3): 905–921. https://doi.org/10.1080/13670050.2020.1730763.

Dylman, A. S., and M. F. Champoux-Larsson. 2020. "It's (Not) All Greek to Me: Boundaries of the Foreign Language Effect." *Cognition* 196:104148. https://doi.org/10.1016/j.cognition.2019.104148.

Fazio, R. H., D. M. Sanbonmatsu, M. C. Powell, and F. R. Kardes. 1986. "On the Automatic Activation of Attitudes." *Journal of Personality and Social Psychology* 50 (2): 229–238. https://doi.org/10.1037/0022-3514.50.2.229.

Gao, S., O. Zika, R. D. Rogers, and G. Thierry. 2015. "Second Language Feedback Abolishes the "hot Hand" Effect During Even-Probability Gambling." *Journal of Neuroscience* 35 (15): 5983–5989. https://doi.org/10.1523/JNEUROSCI.3622-14.2015.

Gawronski, B., J. De Houwer, and J. W. Sherman. 2020. "Twenty-Five Years of Research Using Implicit Measures." *Social Cognition* 38 (Supplement): s1–s25. https://doi.org/10.1521/soco.2020.38.supp.s1.

Geipel, J., C. Hadjichristidis, and L. Surian. 2015. "How Foreign Language Shapes Moral Judgment." *Journal of Experimental Social Psychology* 59:8–17. https://doi.org/10.1016/j.jesp.2015.02.001.

Greene, J. D., S. A. Morelli, K. Lowenberg, L. E. Nystrom, and J. D. Cohen. 2008. "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment." *Cognition* 107 (3): 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (5537): 2105–2108. https://doi.org/10.1126/science.1062872.

Greenwald, A. G., B. A. Nosek, and M. R. Banaji. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology* 85 (2): 197–216. https://doi.org/10.1037/0022-3514.85.2.197.

Hayakawa, S., D. Tannenbaum, A. Costa, J. D. Corey, and B. Keysar. 2017. "Thinking More or Feeling Less? Explaining the Foreign-Language Effect on Moral Judgment." *Psychological Science* 28 (10): 1387–1397. https://doi.org/10.1177/0956797617720944.

Holleman, B., N. Kamoen, and M. 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. https://doi.org/10.1080/02699931.2021.1889472.

Kahneman, D. 2003. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist* 58 (9): 697–720. https://doi.org/10.1037/0003-066X.58.9.697.

Keysar, B., S. L. Hayakawa, and S. G. An. 2012. "The Foreign-Language Effect: Thinking in a Foreign Tongue Reduces Decision Biases." *Psychological Science* 23 (6): 661–668. https://doi.org/10.1177/0956797611432178.

Li, M.-H., P.-W. Li, and L.-L. Rao. 2021. "Self–Other Moral Bias: Evidence from Implicit Measures and the Word-Embedding Association Test." *Personality and Individual Differences* 183:111107. https://doi.org/10.1016/j.paid.2021.111107.

Lieberman, M. D. 2007. "Social cognitive neuroscience: a review of core processes." *Annual Review of Psychology 58* (1): 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654.

Liu, C., L. Li, L. Jiao, and R. Wang. 2021. "Bilingual Language Control Flexibly Adapts to Cultural Context." *Frontiers in Psychology* 12:744289. https://doi.org/10.3389/fpsyg.2021.744289.

Liu, C., H. Wang, K. Timmer, and L. Jiao. 2022. "The Foreign Language Effect on Altruistic Decision Making: Insights from the Framing Effect." *Bilingualism: Language and Cognition* 25 (5): 890–898. https://doi.org/10.1017/S1366728922000128.

Maxwell, S. E., M. Y. Lau, and G. S. Howard. 2015. "Is Psychology Suffering from a Replication Crisis? What Does "Failure to Replicate" Really Mean?" *American Psychologist* 70 (6): 487–498. https://doi.org/10.1037/a0039400.

Muda, R., P. Niszczota, M. Bialek, and P. Conway. 2017. "Reading Dilemmas in a Foreign Language Reduces Both Deontological and Utilitarian Response Tendencies." *Journal of Experimental Psychology: Learning Memory and Cognition* 44 (2): 321–326. https://doi.org/10.1037/xlm0000447.

Nadarevic, L., L. C. Klein, and J. Dierolf. 2021. "Does Foreign Language Alter Moral Judgments? Inconsistent Results from two pre-Registered Studies with the CNI Model." *Open Psychology* 3 (1): 66–86. https://doi.org/10.1515/psych-2020-0112.

Purpuri, S., N. Vasta, R. Filippi, B. Treccani, L. Wei, and C. Mulatti. 2024b. "Beyond the Foreign Language Effect: Unravelling the Impact of L2 Proficiency on Rationality." *Bilingualism: Language and Cognition* 1–8. https://doi.org/10.1017/S1366728924000750.

Purpuri, S., N. Vasta, R. Filippi, L. Wei, and C. Mulatti. 2024a. "Does Language Shape the way we Think? A Review of the Foreign Language Effect Across Domains." *International Journal of Bilingualism*. https://doi.org/10.1177/13670069231225374.

Romero-Rivas, C., R. López-Benítez, and S. Rodríguez-Cuadrado. 2020. "Would you Sacrifice Yourself to Save Five Lives? Processing a Foreign Language Increases the Odds of Self-Sacrifice in Moral Dilemmas." *Psychological Reports* 125 (1): 498–516. https://doi.org/10.1177/0033294120967285.

Stankovic, M., B. Biedermann, and T. Hamamura. 2022. "Not All Bilinguals are the Same: A Meta-Analysis of the Moral Foreign Language Effect." *Brain and Language* 227:105082. https://doi.org/10.1016/j.bandl.2022.105082.

Wei, Y. 2023. "Gender Differences in Moral Judgement and Relevant Factors." *Journal of Education, Humanities and Social Sciences* 22: 569–573. https://doi.org/10.54097/ehss.v22i.13061.

Winskel, H., T. Ratitamkul, V. Brambley, T. Nagarachinda, and S. Tiencharoen. 2016. "Decision-making and the Framing Effect in a Foreign and Native Language." *Journal of Cognitive Psychology* 28 (4): 427–436. https://doi.org/10.1080/20445911.2016.1139583.

Yamaguchi, M., and G. Beattie. 2020. "The Role of Explicit Categorization in the Implicit Association Test." *Journal of Experimental Psychology: General* 149 (5): 809–827. https://doi.org/10.1037/xge0000685.