Bilingualism: Language and Cognition

cambridge.org/bil

Moral identity and language jointly influence how bilinguals feel and decide

Irini Mavrou[1,2,3], Stavroula Mavrou[4] and Andreas Kyriakou[2]

[1]Department of Culture, Communication and Media, University College London, UK; [2]Departamento de Lenguas Aplicadas, Universidad Antonio de Nebrija, Spain; [3]Centro de Investigación Nebrija en Cognición, Universidad Antonio de Nebrija, Spain and [4]Department of Psychology, University of Western Macedonia, Florina, Greece

Research Article

Cite this article: Mavrou, I., Mavrou, S. and Kyriakou, A. (2026). Moral identity and language jointly influence how bilinguals feel and decide. *Bilingualism: Language and Cognition* 1–17. https://doi.org/10.1017/S136672892610159X

Received: 15 August 2025 UTC
Revised: 10 April 2026 UTC
Accepted: 27 May 2026 UTC

Keywords:
moral identity; moral self; moral integrity; moral decision-making; emotionality

Corresponding author:
Irini Mavrou;
Email: i.mavrou@ucl.ac.uk

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

This preregistered study investigated how moral identity relates to moral decisions and emotional responses in 106 Greek–English bilinguals. Participants evaluated two sacrificial-harm and two honesty-based interpersonal moral dilemmas and completed the Moral Identity Questionnaire (Black & Reynolds, 2016, *Personality and Individual Differences*, 97, 120–129). Analyses combined NHST and Bayesian approaches. Higher moral self and moral integrity scores predicted stronger emotional reactions and more deontological choices, particularly in honesty-based dilemmas. Language did not exert a reliable overall effect on moral decisions. However, participants responding in their first language (L1) reported higher moral self scores than those responding in their second language (L2), whereas moral integrity did not differ across language conditions. Emotional intensity varied by language and dilemma type: participants generally reported stronger emotions in their L2 but experienced stronger emotional intensity in response to sacrificial dilemmas in their L1. A multilevel mediation analysis indicated that emotional intensity did not reliably mediate the relationship between language and moral choices.

Highlights

- Moral identity predicts emotional responses and moral decisions across languages.
- Moral self is higher in L1, while moral integrity remains stable across languages.
- Language does not reliably influence moral decisions across dilemma types.
- Emotional intensity is higher in L2 but does not translate into decision differences.
- Emotion does not mediate the relationship between language and moral judgement.

1. Introduction

We live in a complex and highly uncertain world. Contemporary forms of mobility, digitalisation, and global interconnectedness often place individuals in situations requiring them to navigate ethically charged dilemmas. Every day, we are called upon to reflect on what is right and wrong, make decisions, and assume responsibility for their consequences. In many such situations, there are no explicit rules to guide or sanction behaviour, and decision-making often relies on intuition and emotion, on logical reasoning, or most commonly on some interaction between the two (dual-process theory; Greene et al., 2001; Kahneman, 2003). At the same time, globalisation has generated unprecedented levels of multilingual and multicultural contact. Beyond the practical need for communication, increasing numbers of people must now express emotions, evaluate arguments, and make decisions in a second (L2) or foreign language.

In this context, accumulating evidence suggests that the language in which a moral dilemma is presented – first language (L1) versus L2 – can influence bilinguals' moral judgements. The first demonstration of a foreign-language effect in moral dilemmas was reported by Costa et al. (2014), who showed that using an L2 increases utilitarian choices in classic sacrificial moral dilemmas involving harm to one or more individuals for the benefit of others. This empirical pattern, now commonly referred to as the moral foreign-language effect (MFLE), has been replicated across a number of studies (Białek et al., 2019; Dewaele et al., 2024; Geipel et al., 2015a, 2015b; Hayakawa et al., 2017; Kyriakou et al., 2023; Kyriakou & Mavrou, 2025) and has been supported by several recent systematic reviews that assess its robustness and the specific circumstances under which this phenomenon holds (Circi et al., 2021; Del Maschio et al., 2022; Stankovic et al., 2022). These reviews broadly converge in showing that L2 use promotes more utilitarian responding, but they also emphasise the role of moderating factors such as dilemma type and cultural background.[1]

[1]Specifically, Circi et al. (2021) suggested that the magnitude of the MFLE varies as a function of the tasks used and the degree of similarity between L1 and L2. Stankovic et al. (2022) found that the MFLE was moderated by self-reported reading proficiency. In turn, Del Maschio et al. (2022) detected a small but significant foreign language effect across risk and moral-conflict decision-making tasks but reported no significant moderating effects of L2 age of acquisition, proficiency, problem type, task modality, or personal/impersonal framing.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Several explanations have been proposed for the MFLE, although no single account has reached consensus. Specifically, L1 is generally viewed as the language of emotional resonance and of early-acquired moral and social norms (see Caldwell-Harris, 2015, for a review), often associated with more intuitive and deontological decisions (i.e., decisions that prioritise adherence to moral rules or duties). Conversely, L2 has been argued to promote a more deliberative and analytically oriented mode of processing, which tends to favour utilitarian choices (i.e., choices aimed at maximising overall welfare, even when doing so requires endorsing harm to an individual for the benefit of a greater number of people) (Greene et al., 2001; Hayakawa et al., 2017).[2] One influential explanation for this pattern is the 'feeling less (in the L2)' hypothesis, which proposes that reduced emotional resonance in an L2 mediates the relationship between language and moral judgement. Although some studies have provided empirical support for this hypothesis (Dewaele et al., 2024; Kyriakou et al., 2023; but see Geipel et al., 2015b), alternative accounts highlight the inconsistency of this evidence and suggest that the robustness and magnitude of the MFLE depend on linguistic and contextual factors, such as L1–L2 proximity and bilinguals' exposure to the L2 in their L1 environment (Dylman & Champoux-Larsson, 2020; Kyriakou & Mavrou, 2023b; Miozzo et al., 2020).

Recent work further reinforces the importance of such moderators. For example, Privitera et al. (2023) challenged the assumption that using a foreign language reliably dampens emotional responding. Instead, they showed that bilinguals' dominance patterns, immersion experience, and the emotional attachment to L1 versus L2 help determine the impact of language on moral choices. Likewise, Dylman and Champoux-Larsson (2020) argued that the extent to which an L2 induces emotional distancing may be moderated by its role in everyday communication. In their study with Swedish–English bilinguals in Sweden, they found that English, a highly influential and frequently used L2 in the Swedish context, did not elicit a detectable MFLE and suggested that extensive use of English across social and professional domains may reduce the emotional distance associated with its use, thereby attenuating or even eliminating the MFLE (see also Kyriakou & Mavrou, 2023b, for similar findings in Cyprus). In line with these findings, Mavrou et al. (2026) reported no detectable MFLE and attributed its absence to participants' early L2 acquisition and their sustained, everyday engagement with English, conditions likely to reduce emotional resonance in the L2. Related research on the foreign-language effect in other domains, such as framing effects in economic decision-making (Keysar et al., 2012) and cognitive biases including outcome bias and representativeness-based fallacies (Vives et al., 2018), has helped delineate the conditions under which language context may influence decision-making. From a pragmatic perspective, McNab and Mavrou (2025a) further showed that bilinguals evaluated workplace impoliteness more harshly in their L2 than in their L1, despite reporting comparable emotional reactions across languages, suggesting that language context may influence evaluations of interpersonal transgressions independently of self-reported emotional intensity.

Equally important, moral decision-making is not uniform across individuals, since people make decisions depending on their personality traits, motivations, attitudes, previous experiences, and cultural backgrounds (see Warner et al., 2024, for a review). In fact, an emerging line of work underscores the relevance of such individual differences in moral decision-making across languages. For example, Dewaele et al. (2024) reported that higher levels of emotional intelligence were associated with harsher evaluations of offence seriousness and stronger emotional responses to moral transgressions. Further evidence comes from Mavrou et al. (2026), who found that specific executive functions, namely updating and inhibition, predicted bilinguals' moral choices. In a similar vein, McNab and Mavrou (2025b) found that mood and sociability (a facet of trait emotional intelligence) shaped responses to impoliteness, further supporting the view that both stable individual differences and momentary affective states can influence moral-emotional evaluations.

Despite these advances, virtually no research has examined whether more enduring psychological characteristics, such as moral identity, also contribute to variability in bilinguals' moral judgements. An exception is Peressotti et al. (2024), who reported systematic language-related differences in the endorsement of several moral foundations (harm/care, purity, fairness, and ingroup/loyalty), suggesting that language context modulates the relative salience of specific moral values rather than moral judgement overall. However, research has not yet investigated how moral identity, as conceptualised from different perspectives (e.g., moral self and moral integrity in this study), relates to language effects in bilinguals' moral reasoning and emotional responses. This represents a notable gap, given that moral identity is a well-established predictor of ethical behaviour and moral sensitivity (Aquino et al., 2009; Aquino & Reed, 2002; Blasi, 1983; Colby & Damon, 1992; Reynolds & Ceranic, 2007).

In light of these considerations, the current study focused on two components of moral identity – moral self and moral integrity – and was designed to address three goals: (1) to determine whether moral identity assessed in the L1 (Greek) and the L2 (English) reflects the same or distinct underlying constructs; (2) to examine whether moral identity influences moral decisions elicited by two types of moral dilemmas: sacrificial dilemmas, which involve harming one or more individuals for the greater good, and honesty-based dilemmas, which involve interpersonal or self-serving rule violations, with responses interpreted descriptively as endorsing or rejecting the proposed action[3]; and (3) to investigate whether moral identity predicts the emotional responses associated with these dilemmas, as well as whether emotional intensity mediates the link between language and moral choices. To our knowledge, this study is the first to explore the role of moral identity in a bilingual context. The inclusion of these two dilemma types allowed us to examine whether language effects extend beyond classic sacrificial moral dilemmas to situations involving honesty and interpersonal integrity. This design also responds to calls for greater diversity

[2]These utilitarian responses are not necessarily selfish; rather, they reflect a shift toward outcome-based reasoning that may feel less emotionally aversive in an L2 context. Importantly, this characterisation applies primarily to classic sacrificial dilemmas involving explicit trade-offs in harm and benefit and may not generalise to all types of moral scenarios.

[3]In earlier versions of the manuscript and in the preregistration, these dilemmas were described as 'realistic' and 'unrealistic'. Following reviewer feedback, we now refer to them as *sacrificial* and *honesty-based* dilemmas to better capture the moral contexts they represent, rather than implying that the distinction primarily concerns realism. In the present study, the terms sacrificial and sacrificial-harm, as well as honesty-based and honesty-based interpersonal, are used interchangeably.

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

in the types of moral scenarios used in experimental research on moral decision-making, which may help improve the ecological relevance of experimental findings and support broader population-level generalisations (Bauman et al., 2014; Sommer et al., 2010). Studies on deception in bilingual contexts further suggest that L2 use influences willingness to lie (Hu & Navarrete, 2025), indicating that language effects may extend beyond classic sacrificial dilemmas to moral decisions involving honesty and interpersonal integrity, which are directly targeted by the honesty-based dilemmas used in the current study. Moreover, we focused on Greek–English bilinguals, an underrepresented but interesting population from a linguistic point of view, as L1 Greek speakers are typically multilinguals (i.e., they speak at least two additional languages besides their L1[4]), and English has a strong presence in Greek society.[5] The methodology, design, and research hypotheses of this study were preregistered in May 2023 prior to data collection and analysis (https://osf.io/afv5e/).

2. Moral identity and moral judgements

Moral identity is central to moral behaviour and may also help explain individual variability in bilinguals' moral reasoning, a possibility that has not yet been systematically examined in MFLE research. Based on Blasi's framework, Aquino et al. (2009) conceptualised moral identity as 'the cognitive schema a person holds about his or her moral character', which functions as 'a powerful source of moral motivation' (p. 124) (see also Aquino & Reed, 2002). Individuals strive to maintain consistency between their internalised moral schemas and their behaviour; the more central these schemas are to the self, the more likely people are to engage in ethical and prosocial actions. Similarly, Colby and Damon (1992) described moral identity as reflecting a 'unity' between one's moral values and one's sense of self. Black and Reynolds (2016) later conceptualised moral identity as a multidimensional construct comprising moral self (the extent to which moral traits are central to one's self-concept) and moral integrity (the motivation to act in accordance with one's moral self). This distinction highlights that moral identity not only concerns how individuals see themselves but also how strongly they are motivated to translate moral self-views into action. When moral identity is salient and strongly held, it tends to guide behaviour in prosocial and norm-consistent directions. For example, Aquino and Reed (2002) found that moral identity was associated with both self-reported volunteering and actual donation behaviour. Aquino et al. (2007) further showed that individuals with high moral identity were less prone to moral disengagement and less supportive of harsh punitive responses to the perpetrators of the September 11 attacks. Overall, this line of research suggests that moral identity fosters altruistic tendencies and discourages behaviours that harm others or violate moral norms (Aquino et al., 2009).

[4]Recent data by Eurostat (European Commission, 2025) indicate that 35.5% of primary school pupils and over 90% of secondary pupils in Greece learn two or more foreign languages (usually English, in addition to French and German).

[5]According to the 2022 EF English Proficiency Index (EF Education First, 2022), Greece ranks 14th out of 111 countries and regions and falls within the 'high proficiency' category. Tourism, one of Greece's main economic sectors, and the wide availability of English-language television programmes, substantially increase exposure to English as an L2. As a result, English proficiency is a variable that is practically impossible to control for in the Greek context.

3. Moral identity: emotional and contextual dimensions

Moral identity has important affective implications. Because individuals are motivated to act in ways that are consistent with their moral self-concept (see Hardy & Carlo, 2005, for a discussion), emotions are likely to arise when behaviour deviates from internalised moral standards. Blasi (1983) argued that guilt, a key moral emotion, is especially likely when actions contradict one's moral identity. In turn, Stets (2010) emphasised that feelings of guilt or shame emerge when individuals perceive themselves as failing to live up to their standards; guilt in particular motivates reparative behaviours such as apologising and making amends. Importantly, recent work in bilingualism suggests that the emotional experience of moral violations may be modulated by the language in which moral reflection occurs. For example, in Kyriakou et al.'s (2024) study, bilinguals tended to experience guilt less intensely when reflecting on morally transgressive behaviour in their L2 compared to their L1. Their findings indicate that using an L2 can attenuate guilt-related affective responses, particularly in situations involving interpersonal harm or norm-breaking behaviour.

Additional studies reveal that moral identity is malleable and responsive to contextual influences. Goh et al. (2022) demonstrated that prolonged exposure to prosocial messages (e.g., in the form of narrative texts or pictures with embedded text messages) encouraging moral action by targeting specific values (e.g., fairness, kindness, generosity, honesty) heightened moral identity symbolisation, reflecting people's desire to display moral traits. These findings suggest that educational and social contexts can strengthen the salience of moral standards and may therefore also shape the intensity of self-conscious moral emotions, such as guilt or shame, when those standards are violated. This is also relevant because moral identity can influence, and be influenced by, other identity domains such as political, religious, cultural, and even consumer identity (Winterich et al., 2015).

4. Research questions and hypotheses

Against this background, this study examined the extent to which moral identity relates to bilinguals' moral decisions and emotions and was guided by the following research questions:

RQ1. Does moral identity differ when assessed in the L1 (Greek) versus the L2 (English), or do both language versions of the Moral Identity Questionnaire (Black & Reynolds, 2016) tap the same underlying construct?

RQ2. Does moral identity influence Greek–English bilinguals' moral choices and emotional intensity when responding to moral dilemmas presented in either their L1 or their L2?

RQ3. Does emotional intensity mediate the effect of language (L1 versus L2) on these bilinguals' moral choices?

With respect to RQ1, we anticipated that Greek–English bilinguals would show at most small differences in moral identity across language versions of the Moral Identity Questionnaire used in the current study. This expectation reflects the idea that moral identity develops within relatively stable sociocultural frameworks, such as family, cultural norms, and religious tradition, and is anchored in enduring values and practices (Aquino & Reed, 2002; Hardy & Carlo, 2011). In the Greek context, for instance, religion continues to play a prominent role in society, influencing individuals' beliefs, behaviours, and social identities (Sakellariou, 2013). Although

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

these sociocultural factors were not directly measured in the present study, they provided a theoretical rationale for treating moral identity as comparable across language conditions as a plausible starting point, while remaining open to potential language-related differences. Additional support for this hypothesis comes from Black and Reynolds (2016), who conceptualise moral identity as a relatively stable aspect of self-concept. From this perspective, substantial language-dependent shifts in moral identity would not be expected.

Regarding RQ2, and drawing on prior research (Aquino et al., 2009; Aquino & Reed, 2002; Hardy & Carlo, 2011), we expected moral identity to predict both moral decisions and emotional intensity across languages. Previous work has consistently shown that individuals for whom moral identity is highly central to the self exhibit greater prosocial motivation, stronger adherence to moral norms, and more harm-averse responses (Aquino et al., 2007; Aquino & Reed, 2002; Blasi, 1983; Hardy & Carlo, 2011). However, commitment to moral values does not necessarily prescribe a single decision strategy; rather, strongly internalised moral standards may give rise to different patterns of moral decision-making across contexts. Consistent with our preregistration, we anticipated that higher moral identity would be associated with more deontological decisions while recognising that such decisions reflect a tendency to prioritise rule-based considerations rather than the absence of alternative forms of moral commitment. Moreover, we acknowledge that outcome-based (i.e., consequentialist) choices may also, in some contexts, reflect strong moral commitment. Accordingly, we predicted that higher moral identity would be associated with moral choices reflecting stronger adherence to one's moral standards, together with heightened emotional intensity.

Lastly, for RQ3, and in line with accounts of reduced emotional resonance in the L2, we hypothesised that any influence of language on moral choices could operate indirectly through emotional responses (Dewaele et al., 2024; Kyriakou et al., 2023). In other words, although we did not expect language to meaningfully alter moral identity, we anticipated that it might influence the emotional intensity associated with moral decision-making, with downstream consequences for moral decision outcomes.

5. Method

5.1. Participants

One hundred and fifty-one participants were initially recruited from different universities in Greece, social media platforms, and personal contacts using snowball sampling. However, the data from 45 of them were incomplete and were therefore discarded[6], resulting in a final sample of 106 L1 Greek–L2 English bilinguals: 20 males and 85 females (one participant did not report their gender), aged between 18 and 55 years, and distributed across the following age groups: 18–25 years old (n = 48), 25–35 years old (n = 33), 36–45 years old (n = 13), and 46–55 (n = 11; one participant did not report their age). All participants had an upper-intermediate or advanced level of English, as assessed using 7-point Likert scales measuring L2 English proficiency in reading, writing, speaking, and listening (M = 5.35, SD = 1.16, N = 106) and the Lexical Test for Advanced Learners of English (Lemhöfer & Broersma, 2012), a lexical decision task that provides a satisfactory estimate of vocabulary size in English, with scores above 60 indicating an upper-intermediate B2 level.[7] Fifty-four participants (females = 41) completed the tasks in their L1 Greek, while the remaining 52 did so in their L2 English, and none of them knew in advance which version of the tasks they would be asked to complete.

5.2. Materials

5.2.1. Moral dilemmas

We used two sacrificial-harm moral dilemmas (the vaccine test dilemma and the transplant dilemma) and two honesty-based interpersonal dilemmas (the resume dilemma and the cheater's dilemma) drawn from Koenigs et al. (2007) and Kyriakou and Mavrou (2023a). In the vaccine test dilemma, participants had to decide whether they would kill one person with a deadly injection in order to identify a vaccine capable of saving millions of lives. In the transplant dilemma, they had to decide whether to transplant a young man's organs against his will into five patients, thereby resulting in his death. In the resume dilemma, they were asked to choose whether to include false information in their CV to increase their chances of being hired. In the cheater's dilemma, they had to decide whether to disclose an act of infidelity to their partner. Because the original materials were written in English, forward and back translations were conducted by two Greek–English bilinguals who were licensed psychologists and had experience with psychometric validity assessment.

It is important to clarify that the distinction between sacrificial and honesty-based dilemmas does not represent a theoretically pure dichotomy, but rather a practical way to differentiate classic sacrificial-harm dilemmas commonly used in moral psychology from dilemmas involving honesty-related interpersonal decisions. In fact, moral dilemmas vary along multiple psychological dimensions (e.g., likelihood of occurrence, emotional arousal, personal involvement, conflict intensity), rather than constituting discrete psychological categories (Christensen et al., 2014; Kahane et al., 2015). In our study, sacrificial-harm dilemmas involve directly causing harm for the sake of aggregate welfare, whereas honesty-based dilemmas centre on interpersonal decisions involving deception or disclosure without physical harm. The present design therefore contrasts two moral contexts rather than manipulating realism as a single psychological dimension.

The materials were selected both to align with established moral psychology paradigms and to extend research on the MFLE beyond classic sacrificial scenarios, as explained earlier. Sacrificial moral dilemmas, including variants such as vaccine and transplant scenarios, have been widely used in previous research, and MFLE patterns have been repeatedly observed with these paradigms. In contrast, the resume and cheater's dilemmas require weighing honesty-based norms against personal or relational motives and fall within the broader category of honesty- or deception-related dilemmas, which involve deception or interpersonal obligations. This distinction also connects them to research showing that foreign-language effects can emerge in moral decisions involving honesty or deception (e.g., Hu & Navarrete, 2025; Kyriakou & Mavrou, 2023b).

[6]These participants opened the online questionnaire and responded to some background questions but did not read or respond to the moral dilemmas, nor did they complete the Moral Identity Questionnaire.

[7]According to our preregistration, data from participants scoring below 60 on the LexTALE were to be excluded. Fifteen participants scored slightly below this cut-off, although all scored above 50; therefore, these data were retained for the subsequent analyses. In addition, six participants did not complete the LexTALE but self-reported medium to high L2 English proficiency, and their data were likewise retained because they completed the Moral Identity Questionnaire.

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

Across the four dilemmas, the sacrificial-harm scenarios contrasted deontological (rule-based) and utilitarian (outcome-based) responses and varied in conflict severity. The vaccine test dilemma represents a particularly high-conflict sacrificial scenario, as the potential benefit extends to millions of people, which may increase utilitarian responding in both L1 and L2. Still, individuals with strong moral commitments may evaluate such dilemmas differently depending on the moral principles they prioritise, which may include both rule-based and outcome-based considerations, rather than reflecting a single type of moral reasoning. By contrast, the honesty-based interpersonal dilemmas required participants to choose between being transparent by telling the truth or acting in self-interest without causing bodily harm or legal consequences (e.g., lying on a resume). In both dilemmas, refusing to lie was treated as the deontological option. Although honesty is not always 'morally superior' and lies may sometimes be justified, the resume dilemma primarily involves deception for personal gain, whereas in the cheater's dilemma, lying may aim to protect oneself or a partner from emotional harm. Importantly, endorsement of the action in these dilemmas is interpreted descriptively (i.e., as action endorsement or rule violation) rather than as strictly utilitarian responding, as these scenarios do not necessarily involve explicit welfare maximisation. The key difference between these dilemmas is that the immoral behaviour has already occurred in the cheater's scenario, while in the resume scenario it concerns a planned action. Despite this difference, both types of honesty-based dilemmas represent plausible moral conflicts that even young adults can meaningfully evaluate without relying on direct personal experience (Haidt, 2001). Additionally, dilemmas involving workplace dishonesty (e.g., lying, deceit, and subterfuge at work; Grover, 1993), as well as honesty-related conflicts in romantic relationships (Sommer et al., 2010), are well established in the literature and can be regarded as appropriate for a wide age range.[8]

5.2.2. Moral identity questionnaire

Participants' moral identity was assessed using the Moral Identity Questionnaire (MIQ; Black & Reynolds, 2016), which comprises 20 statements rated on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree). These items assess two established components of moral identity, both of which were examined in separate statistical models. Eight items assess moral self (MIQ-MS), defined as 'the importance people place on being moral' (*p.* 121), and 12 items assess moral integrity (MIQ-MI), referring to 'the desire to make intention and action consistent, and how much value participants place on acting according to moral principles' (*p.* 121). Higher scores indicate a greater perceived importance of moral self and integrity, respectively. Black and Reynolds (2016) reported satisfactory psychometric properties for the original instrument (α = .84 for MIQ-MS; α = .87 for MIQ-MI; test–retest reliability *r*tt = .87).

To our knowledge, no standardised Greek version of the MIQ is currently available. Therefore, we followed standard cross-linguistic adaptation procedures to ensure conceptual and linguistic equivalence across languages. The items were initially translated into Greek by a bilingual co-author and independently back-translated by another bilingual co-author, both of whom are trained psychologists. Discrepancies were discussed and resolved collaboratively to preserve semantic accuracy. The resulting Greek version was then reviewed to ensure clarity, cultural appropriateness, and alignment with the original English version.

Internal reliability analyses indicated acceptable consistency for both subscales in the present sample, across both language versions of the MIQ (Greek version: Cronbach's α = .813 for MIQ-MI and .660 for MIQ-MS; English version: Cronbach's α = .768 for MIQ-MI and .656 for MIQ-MS). Although the MIQ-MS subscale yielded somewhat lower internal consistency than the MIQ-MI subscale, these values fall within the range commonly reported for moral identity measures and are considered acceptable for research purposes (see, e.g., Aquino & Reed, 2002). Moreover, Cronbach's alpha is known to be sensitive to scale length and item heterogeneity; shorter scales or those assessing less internally homogeneous constructs frequently yield alpha coefficients in the mid-60s while still demonstrating adequate construct validity (Cortina, 1993; Schmitt, 1996; Streiner, 2003). This issue is further amplified in translated or culturally adapted instruments, where linguistic and cultural variation in item interpretation can attenuate internal consistency without necessarily compromising the underlying construct being measured (e.g., Hambleton et al., 2005; Van de Vijver & Hambleton, 1996). On this basis, item removal was not pursued, as retaining the full scale preserved the established two-factor structure of the MIQ and maintained comparability with prior research. Moreover, because MIQ-MS and MIQ-MI were entered in separate statistical models, any attenuation associated with the lower reliability of the MIQ-MS subscale could not bias the results involving MIQ-MI or the language predictor. If anything, reduced reliability would be expected to attenuate observed effects, rendering the present findings conservative rather than inflated.

5.3. Procedure

Participants were recruited through university mailing lists, social media platforms, and personal contacts, and voluntarily enrolled after providing informed consent. All participants met the eligibility criteria of being L1 Greek speakers with at least an intermediate level of English proficiency. Participants were allocated to language conditions based on the version of the survey they accessed (Greek or bilingual Greek–English), which were distributed in parallel during the recruitment period. The materials were presented via the Qualtrics survey platform (Qualtrics, 2022). In the L1 (Greek) condition, all study materials, including the consent form, instructions, moral dilemmas, emotional intensity ratings, and the MIQ, were presented entirely in Greek. In the L2 (English) condition, moral dilemmas and the MIQ were presented in English, while administrative components (e.g., the consent form, general instructions) were provided in Greek to ensure full comprehension of the ethical and procedural information. This constitutes a partial language-immersion design, which we note as a methodological limitation; however, the language manipulation was applied consistently to all task-relevant components in the L2 condition.

After providing informed consent, participants read each dilemma, indicated on a 7-point Likert-type scale the extent to which the dilemma evoked strong emotions, and then responded to the moral question accompanying each dilemma using a dichotomous yes/no scale. For each dilemma, the moral question asked participants whether they would personally carry out the described action (e.g., 'Would you include false information in your CV?', 'Would you tell your partner that you cheated on them?'). A 'yes' response indicated endorsement of the action, and a 'no' response indicated rejection of the action. The order of the four dilemmas

[8]To ensure that our conclusions were not driven by the specific features of any single scenario, we additionally conducted analyses treating each dilemma separately (see Appendix 3B of the Supplementary Material).

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

was randomised across participants using the Qualtrics randomisation function. Emotional intensity was measured before the moral judgement question in order to preserve the temporal ordering assumed in mediation models and to minimise the possibility that participants' explicit decisions would influence their emotional ratings. This is consistent with recommendations that the mediator should be measured prior to the outcome in causal mediation models (Tingley et al., 2013).[9] Participants then completed the MIQ and the LexTALE and responded to questions about their demographic and language backgrounds.

The study was approved by the Research Ethics Committee of Nebrija University (reference number: UNNE-2023-00011) and the Research Ethics Committee of the University of Western Macedonia (reference number: 225/2023) and was conducted in accordance with the Declaration of Helsinki and the ethical guidelines of the American Psychological Association. All study materials, including the original English dilemmas, Greek translations, the MIQ, and all accompanying instructions, are available on the Open Science Framework (OSF), together with the anonymised dataset and analysis code (https://osf.io/emvaz/).

5.4. Data analysis

The study hypotheses, design, and analysis plan were preregistered on OSF. The preregistration included an *a priori* power analysis informed by the analytical plan at the time. This calculation assumed a simplified regression model with three predictors: language, dilemma type, and moral identity. Using G*Power (Faul et al., 2007, 2009), we estimated that a total sample of $N = 77$ would be sufficient to detect a medium effect size ($f^2 = .15$) with $\alpha = .05$ and power = .80. To allow for attrition and unusable data, we preregistered a target of up to 100 participants, which we exceeded (final $N = 106$, after excluding 45 participants). However, the mixed-effects models ultimately involved additional complexity not accommodated in the preregistered power analysis. Accordingly, we implemented several safeguards to ensure model stability: (a) retaining random intercepts for both participants and dilemmas in all models; in some models, this specification resulted in singular fits, with variance estimates for the participant random effect approaching zero, indicating minimal residual between-participant variability; nevertheless, the full random-effects structure was retained to preserve the theoretically motivated model specification; (b) conducting complementary stimulus-level analyses in which the four dilemmas were entered as separate fixed levels; and (c) comparing results from the full sample with those from a more homogeneous age-restricted subgroup. Convergent patterns across these analyses suggest that the main effects were not unduly driven by model specification or sample size limitations.

Regarding age, and as will be reported in the Results section, there was a significant age difference between the two language conditions. Although preliminary Bayesian analyses suggested that age was unlikely to confound the comparisons between language conditions, we conducted supplementary analyses on a more homogeneous subgroup of participants aged 18–39 years, who represented the largest proportion of the sample ($n = 90$). Our main analyses were conducted on the full sample, while the subgroup of participants aged 18–39 served as a robustness check to assess whether a small number of older participants unduly influenced the overall pattern. The results from this subgroup closely mirrored those of the full sample. Given our sample size and because age was not central to our hypotheses, we did not include age as a covariate in the mixed-effects models to avoid overparameterisation.

To address the first research question concerning potential language-related differences in moral identity (RQ1), we first ran independent-samples *t*-tests comparing MIQ-MS and MIQ-MI scores across the L1 and L2 conditions, using both NHST and Bayesian approaches in JASP (JASP Team, 2024). The Bayesian analyses were included to quantify evidence both for and against language-related differences in moral identity, given our expectation that any such differences would be small. Bayesian analyses were conducted using the default priors implemented in JASP. For Bayesian independent-samples *t*-tests, the prior on the standardised effect size δ was a Cauchy distribution centred at zero with scale parameter $r = .707$ (Rouder et al., 2009; Rouder & Morey, 2012).[10]

To answer the second research question referring to the influence of moral identity on moral judgements and emotions (RQ2), we ran four mixed-effects regression models in RStudio 2022.12.0 (Posit Team, 2022) using the *glmer* function in the *lme4* package (Bates et al., 2015). Two models tested the impact of moral self on moral decisions and emotional intensity, respectively, and two parallel models examined the effects of moral integrity on the same outcome variables. All models were fitted first to the full sample and then re-estimated in the subgroup of participants aged 18–39 years to assess robustness. Before fitting these models, we followed a theory-driven model-building strategy. All main fixed effects (i.e., language condition, dilemma type, and moral identity scores) were specified *a priori* in the preregistration. Dilemma type refers to the contrast between sacrificial-harm dilemmas and honesty-based interpersonal dilemmas. The interaction between language condition and dilemma type was included to test theoretically motivated moderation effects, rather than as part of a stepwise or data-driven selection procedure. This approach helped ensure conceptual coherence across models and reduce the risk of overfitting. Because each participant responded to multiple dilemmas, all mixed-effects models specified random intercepts for participants and dilemmas to account for within-participant dependence and stimulus-level variability, respectively. In the moral decision models, the main preregistered analyses included dilemma type as a fixed predictor together with random intercepts for participants and dilemmas. Additionally, complementary analyses treated the four dilemmas as separate fixed levels rather than as a random effect. This approach was adopted because the study employed only four dilemmas, and the moral decision models including both participant and dilemma random intercepts resulted in singular fits, suggesting that the variance associated with the participant random effect was negligible, although the full random-effects structure was retained. This modelling strategy allowed us to account for repeated measures while avoiding overparameterisation of the random-effects structure. In all models, categorical predictors were coded using

[9]Although many previous studies have presented the moral question before the emotional intensity rating, our preregistered mediation framework required reversing this order. Importantly, the emotional intensity question involved only an immediate affective evaluation and did not require participants to justify or reflect on their decisions. Moreover, the distribution of decisions closely mirrored those reported in prior L1/L2 studies using the standard ordering, suggesting that this design choice did not meaningfully affect the results.

[10]Bayesian analyses were used selectively. For RQ1, their use was appropriate because the null hypothesis was theoretically motivated. In contrast, the mixed-effects models used to address RQ2 were analysed within a frequentist framework due to model complexity and considerations of computational feasibility.

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

treatment (dummy) coding. For language condition, 0 indicated L1 (Greek) and 1 indicated L2 (English). For dilemma type, 0 corresponded to sacrificial-harm dilemmas and 1 to honesty-based interpersonal dilemmas. Moral identity scores (MIQ-MS and MIQ-MI) were entered as continuous, mean-centred predictors.

With regard to emotional intensity ratings, these were collected using 7-point Likert scales and were therefore ordinal in nature but were modelled as continuous variables in the main analyses. Treating Likert-type data with five or more response categories as approximately interval-scaled is widely considered acceptable when distributions are not severely skewed and when the focus is on estimating effects within complex models (Norman, 2010; Rhemtulla et al., 2012). This approach also facilitates the interpretation of interaction effects in mixed-effects models. Nonetheless, to ensure that the results were not dependent on this assumption, we additionally conducted cumulative link mixed models (CLMMs). These ordinal models, which are reported in the Results section, yielded a highly similar pattern of results, supporting the robustness of the overall findings.

Furthermore, although dilemma type was treated as a categorical predictor in the main analyses, we recognised that the four dilemmas varied in content and structure. Broader category-level contrasts based on two dilemmas per category may obscure scenario-specific variation. Therefore, we conducted complementary mixed-effects models in which the four dilemmas were entered as separate fixed levels. These analyses allowed us to assess whether the observed effects reflected the broader dilemma-type contrast or were driven by particular scenarios. The corresponding results are reported in Appendix 3A of the Supplementary Material. Additional dilemma-specific regression analyses are reported in Appendix 3B as a further robustness check. The convergence between the main analyses, the four-dilemma complementary models, and the dilemma-specific supplementary analyses suggests that the observed patterns were not driven solely by any single scenario. Nevertheless, conclusions regarding differences between dilemma types should be interpreted with appropriate caution.

A further consideration relates to L2 proficiency, indexed by LexTALE scores. These scores were not included as a covariate because the study's preregistered hypotheses centred on moral identity rather than variation in L2 proficiency. Moreover, English proficiency was relatively homogeneous in the sample; therefore, including it as a covariate would have offered limited explanatory power and increased the risk of overfitting the statistical models. For these reasons, proficiency was not modelled as a fixed-effect predictor.

To address the third research question (RQ3), we examined whether emotional intensity mediated the relationship between language condition and moral choices using a multilevel mediation approach. Since each participant provided multiple responses across dilemmas, both the mediator and outcome models were estimated as mixed-effects models including random intercepts for participants. The mediator model predicted emotional intensity from language condition and dilemma, whereas the outcome model predicted moral decision (0 = rejection of the action, 1 = action endorsement) from language condition, emotional intensity, and dilemma. These models were estimated using the *lme4* package, and the indirect (Average Causal Mediation Effect; ACME) and direct (Average Direct Effect; ADE) effects were obtained using the *mediation* function from the *mediation* package (Tingley et al., 2014), which implements quasi-Bayesian simulation. The analysis was based on 1,000 simulations to obtain stable confidence intervals for the mediation estimates. This modelling strategy allowed us to account for the nested structure of the data while estimating the indirect and direct effects of language on moral decision through emotional intensity.

6. Results

6.1. Preliminary analyses: Age and L2 proficiency differences between language groups

Comparisons between participants' age and L2 English proficiency across language conditions are summarised in Table 1. Bayes factors for these comparisons are also included. BF_{10} values between 1 and 3 indicate weak evidence for the alternative hypothesis, values between 3 and 10 moderate evidence, and values above 10 strong evidence. BF_{01} values are interpreted analogously as evidence in favour of the null hypothesis (Lee & Wagenmakers, 2013). Overall, L2 English proficiency was broadly comparable across the two language groups. However, participants in the L1 condition were significantly older than those in the L2 condition; therefore, we conducted additional analyses to examine whether age was related to any of the dependent measures. Bayesian correlation tests showed moderate evidence for no association between age and emotional intensity ($r = -.09$, $BF_{01} = 3.39$), suggesting that age was not systematically associated with participants' emotional responses. In addition, a Bayesian logistic regression with decision (0 = rejection of the action, 1 = action endorsement) as the dependent variable and language condition and age as predictors showed that models including age were not preferred over the null model (all $BF_{10} \leq 0.65$, $R^2 \approx .01$), providing further evidence that age did not reliably predict moral decisions in this study.

We repeated the above Bayesian analyses in a subset of participants aged 18–39, who constituted the largest proportion of the sample ($n = 90$). For this subgroup, the Bayesian correlation between age and emotional intensity remained very small and negative ($r = -.11$, $BF_{10} = 0.62$), providing anecdotal evidence in favour of the null hypothesis and again suggesting that age did not meaningfully relate to emotional responses. Likewise, a Bayesian logistic regression restricted to participants aged 18–39 showed that models including age were not preferred over the null model (all $BF_{10} < 1$, $R^2 \approx .01$). This supports the conclusion that age was unlikely to confound the comparisons between language conditions.

6.2. Language differences in moral identity

Descriptive statistics for moral decisions across language conditions and dilemma types are presented in Table 2. Additional descriptive statistics are provided in Appendix 1 of the Supplementary Material, including distributions of emotional intensity ratings by language and dilemma type, as well as summaries for both the full sample and the subgroup of participants aged 18–39. Appendix 2 of the Supplementary Material displays descriptive plots illustrating the relationships among the main variables.

First, we examined differences in MIQ scores across language conditions (see Table 3). The independent-samples *t*-tests revealed a significant effect of language on MIQ-MS scores, such that participants in the L1 (Greek) condition reported higher moral self scores than those in the L2 (English) condition. Cohen's *d* indicated a medium effect size, and the Bayesian analysis provided strong evidence in favour of the alternative hypothesis. In contrast, moral

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

Table 1. Participants' age and L2 English proficiency by language condition (L1 versus L2)

	L1 Greek *M* (*SD*)	L2 English *M* (*SD*)	*t*	*p*	BF_{10}
Age	33.57 (9.92)	24.42 (7.28)	10.730	< .001	4.94×10^{20}
Self-reported L2 reading	5.67 (1.09)	5.75 (0.76)	−0.916	.360	0.161
Self-reported L2 writing	5.20 (1.39)	4.88 (1.16)	2.566	.011	2.498
Self-reported L2 speaking	5.15 (1.26)	5.10 (1.13)	0.448	.655	0.119
Self-reported L2 listening	5.56 (1.26)	5.50 (1.14)	0.477	.634	0.120
Self-reported overall L2 proficiency	5.39 (1.13)	5.31 (0.88)	0.870	.385	0.160
LexTALE	70.31 (11.87)	72.01 (10.74)	−1.497	.135	0.323

integrity was not affected by the language used, with moderate evidence in favour of the null hypothesis.

In addition to reporting Bayes factors from standard Bayesian *t*-tests, we also conducted Bayesian sequential analyses, which plot the Bayes factor as data accumulate across participants. These sequential plots show how the strength of evidence (BF_{10}) stabilises with increasing sample size. Bayesian sequential analyses for moral self and moral integrity are presented in Figures 1 and 2, respectively. Figure 1 shows strong and stable evidence for a language effect on moral self, whereas Figure 2 indicates consistent support for no difference in moral integrity across language conditions. Taken together, these results provide partial support for our first hypothesis: moral self differed significantly across language conditions, whereas moral integrity showed no evidence of a language effect.

6.3. Influence of moral identity on moral decisions and emotional intensity

We ran mixed-effects models to examine the role of moral identity in moral decisions and emotional responses (Table 4). Moral self significantly predicted moral decisions ($B = -0.462$, $p = .018$), and moral integrity showed an even stronger association with decision outcomes ($B = -0.544$, $p < .001$). These findings indicate that participants with higher moral identity were less likely to endorse the described action. In addition, moral self was positively associated with emotional intensity ($B = 0.675$, $p = .005$), suggesting that individuals with stronger moral self-concepts experienced stronger emotional reactions when evaluating the dilemmas. We estimated the same mixed-effects models for the subgroup of participants aged 18–39, and the pattern of results for moral identity mirrored that observed in the full sample (see Appendix 5 in the Supplementary Material).

Language condition did not exert a statistically significant direct effect on moral decisions in the mixed-effects models, and separate analyses for each dilemma confirmed this result. In contrast, dilemma type significantly predicted moral choices in both decision models (Model 1 and Model 2 in Table 4), with participants being more likely to endorse the action in honesty-based dilemmas than in sacrificial dilemmas (Model 1: $B = 1.442$, $p = .038$; Model 2: $B = 1.481$, $p = .036$). The interaction between language condition and dilemma type was not statistically significant in either model.

Since each dilemma category was represented by only two scenarios, we also estimated complementary mixed-effects logistic models in which the four dilemmas were entered as separate fixed levels while retaining participant random intercepts (see Supplementary Material, Appendix 3A). These analyses revealed significant differences across specific dilemmas, with some scenarios eliciting more frequent action endorsement than others. Importantly, however, the overall pattern remained consistent with the main models: moral self ($B = -0.470$, $p = .017$) and moral integrity ($B = -0.558$, $p < .001$) continued to predict a greater tendency to reject the action (i.e., deontological responding), whereas language condition did not emerge as a reliable predictor of moral judgements in these models. Interactions between language condition and individual dilemmas were not statistically significant. These results indicate that the main findings were not driven solely by aggregating dilemmas into two dilemma types and remained robust when the dilemmas were modelled separately.

With respect to emotional responses, a significant interaction between language and dilemma type emerged in both models assessing emotional intensity (Model 3: $B = 0.648$, $p = .019$; Model 4: $B = 0.648$, $p = .018$). Participants in the L2 condition reported similar levels of emotional intensity for sacrificial and honesty-based dilemmas, and their ratings were generally higher than those of participants in the L1 condition. In contrast, participants in the L1 condition reported stronger emotions in response to sacrificial-harm dilemmas than to honesty-based dilemmas. Figure 3 illustrates these patterns by displaying emotional intensity separately for each dilemma type within each language condition.

Table 2. Moral decisions by language condition and dilemma type

Language	Dilemma type	Deontological	Action endorsement	Total
L1	Sacrificial	93	15	108
L1	Honesty-based	70	38	108
Total L1	—	163	53	216
L2	Sacrificial	79	25	104
L2	Honesty-based	64	40	104
Total L2	—	143	65	208
Grand total	Sacrificial	172	40	212
Grand total	Honesty-based	134	78	212

Note: Values represent raw counts of participants' 'Yes/No' moral decisions. 'Deontological' refers to rejecting the action, which in sacrificial dilemmas corresponds to adherence to harm-avoidance rules. 'Action endorsement' refers to endorsing the described action. This table aggregates decisions across two sacrificial-harm dilemmas (vaccine and transplant) and two honesty-based interpersonal dilemmas (resume and cheater). The total column refers to the total number of decisions made within each combination of language condition and dilemma type. The grand total indicates the number of decisions aggregated across language conditions for each dilemma type.

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

Table 3. Moral identity scores across language conditions

	L1	L2	*t*	*p*	Cohen's *d*	BF_{10}	*t controlling for age*	Cohen's *d controlling for age*	BF_{10} *controlling for age*
	M (*SD*)	*M* (*SD*)							
Moral self (MIQ-MS)	5.218 (0.527)	4.950 (0.614)	4.833	< .001	0.470	6,778.60	4.006	0.421	229.632
Moral integrity (MIQ-MI)	3.820 (0.811)	3.728 (0.694)	1.250	.212	0.121	0.229	1.259	0.132	0.249

Note: 'Controlling for age' refers to analyses in which age was included as a covariate in the comparison between language groups.

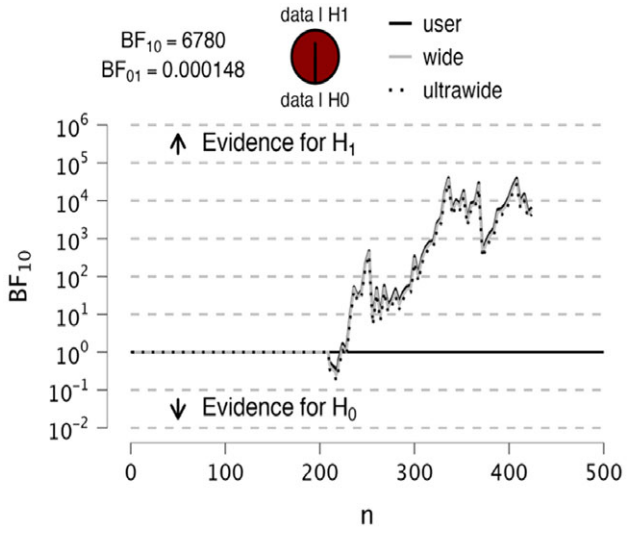


Figure 1. Sequential analysis for moral self. *Note:* The plot illustrates how the Bayes factor (BF_{10}) evolves as each participant is added to the analysis. The steadily increasing BF_{10} indicates growing evidence in favour of a difference in MIQ-MS scores between the L1 and L2 conditions.

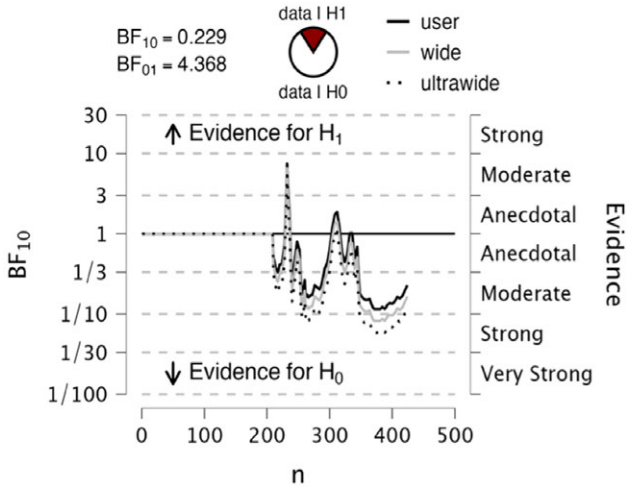


Figure 2. Sequential analysis for moral integrity. *Note:* The Bayes factor (BF_{10}) remains near or below 1 as data accumulate, indicating no evidence in favour of a difference in MIQ-MI scores between the L1 and L2 conditions.

Because emotional intensity was measured on 7-point Likert scales, we conducted additional analyses using cumulative link mixed models (CLMMs) implemented in the *ordinal* package in R (Christensen, 2023) to account for the ordinal structure of the response variable. For both the moral self and moral integrity models, the CLMMs included the same fixed-effects structure as the linear mixed-effects models (language condition, dilemma type, and the corresponding moral identity predictor) and random intercepts for participants and dilemmas. All models converged successfully and yielded a highly similar pattern of results. The interaction between language and dilemma type remained statistically significant in both analyses (moral integrity: $p = .034$; moral self: $p = .035$), mirroring the findings from the linear models. The effects of moral identity were also consistent across model specifications, with MIQ-MS scores continuing to predict higher emotional intensity ($p = .006$) and MIQ-MI scores showing a marginal trend ($p = .099$), in line with the linear mixed-effects results. The consistency of effects across models suggests that the observed patterns in emotional intensity are not contingent on treating Likert-scale responses as interval data but persist under alternative analytic specifications.

Finally, to further illustrate how these effects manifested across individual scenarios, we conducted separate regression analyses for each dilemma (reported in Appendix 3B of the Supplementary Material). These analyses provided additional insight into the observed relationships. Moral self significantly predicted moral decisions in the cheater's dilemma, and the effect of moral integrity approached significance in that scenario. In addition, moral integrity significantly predicted moral choices in the resume dilemma. Furthermore, moral self predicted emotional intensity in three of the four dilemmas (vaccine, transplant, and cheater), whereas moral integrity was linked to emotional responses in the cheater's dilemma. Additional analyses indicated that language influenced emotional responses in three of the four dilemmas (vaccine, resume, and cheater). Participants in the L2 condition reported stronger emotional intensity in these dilemmas than those in the L1 condition, and these differences were statistically significant (see Appendix 4 of the Supplementary Material).

6.4. Mediation of language effects on moral decisions via emotion

Building on theoretical accounts proposing that emotional reactions may influence moral decisions, we next addressed our third research question: whether the effect of language on moral choices operates indirectly through emotional intensity. As each participant evaluated multiple dilemmas, we conducted a multilevel mediation analysis to account for the repeated-measures structure of the data. Specifically, emotional intensity was modelled as a function of language condition and dilemma using a mixed-effects model with random intercepts for participants. Moral decisions were then predicted from language condition, emotional intensity, and dilemma using a mixed-effects logistic regression model that also included participant-level random intercepts. This modelling strategy ensured that the non-independence of repeated responses

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

Table 4. Decision and emotional intensity as a function of language condition, dilemma type, and moral identity

	Fixed effects	B	SE	z/t	Pr(>|z|)
Model 1:	(Intercept)	0.342	1.125	0.304	.761
Decision and	Language (English)	0.609	0.379	1.605	.109
MIQ-MS	Dilemma type (Honesty-based)	1.442	0.694	2.077	.038*
	MIQ-MS	−0.462	0.195	−2.362	.018*
	Language*Dilemma type	−0.592	0.474	−1.247	.212
Random effects	Variance	*SD*			
Participant intercept	≈ 0	≈ 0			
Dilemma intercept	0.344	0.586			
R^2 marginal	.125				
Model 2:	(Intercept)	−0.038	0.749	−0.050	.959
Decision and	Language (English)	0.718	0.380	1.888	.059
MIQ-MI	Dilemma type (Honesty-based)	1.481	0.708	2.091	.036*
	MIQ-MI	−0.544	0.150	−3.621	< .001***
	Language*Dilemma type	−0.615	0.479	−1.281	.200
Random effects	Variance	*SD*			
Participant intercept	≈ 0	≈ 0			
Dilemma intercept	0.359	0.599			
R^2 marginal	.153				
Model 3:	(Intercept)	0.808	1.337	0.644	.521
Emotion and	Language (English)	0.636	0.309	2.057	.041*
MIQ-MS	Dilemma type (Honesty-based)	−0.898	0.686	−1.309	.312
	MIQ-MS	0.675	0.236	2.854	.005**
	Language*Dilemma type	0.648	0.274	2.358	.019*
Random effects	Variance	*SD*			
Participant intercept	1.428	1.195			
Dilemma intercept	0.433	0.658			
R^2 conditional	.532				
R^2 marginal	.096				
Model 4:	(Intercept)	3.003	0.866	3.468	.002***
Emotion and	Language (English)	0.487	0.308	1.580	.116
MIQ-MI	Dilemma type (Honesty-based)	−0.898	0.686	−1.309	.312
	MIQ-MI	0.348	0.182	1.906	.059
	Language*Dilemma type	0.648	0.274	2.358	.018*
Random effects	Variance	*SD*			
Participant intercept	1.510	1.229			
Dilemma intercept	0.433	0.658			
R^2 conditional	.532				
R^2 marginal	.078				

Note: *p < .05, **p < .01, ***p < .001. MIQ-MS: Moral self. MIQ-MI: Moral integrity. Dilemma type contrasts sacrificial-harm dilemmas (vaccine, transplant) with honesty-based interpersonal dilemmas (resume, cheater). The estimated variance of the participant random intercept in Models 1 and 2 was extremely small, indicating minimal residual between-participant variability after accounting for the fixed effects.

within participants was appropriately accounted for in both the mediator and outcome models. The indirect and direct effects were subsequently estimated using the *mediate* function with quasi-Bayesian simulations (1,000 draws). It should also be noted that, although moral identity was entered as a predictor in the mixed-effects models reported earlier, it was omitted from the mediation models, which were specified to test the indirect effect of language via emotional intensity.

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

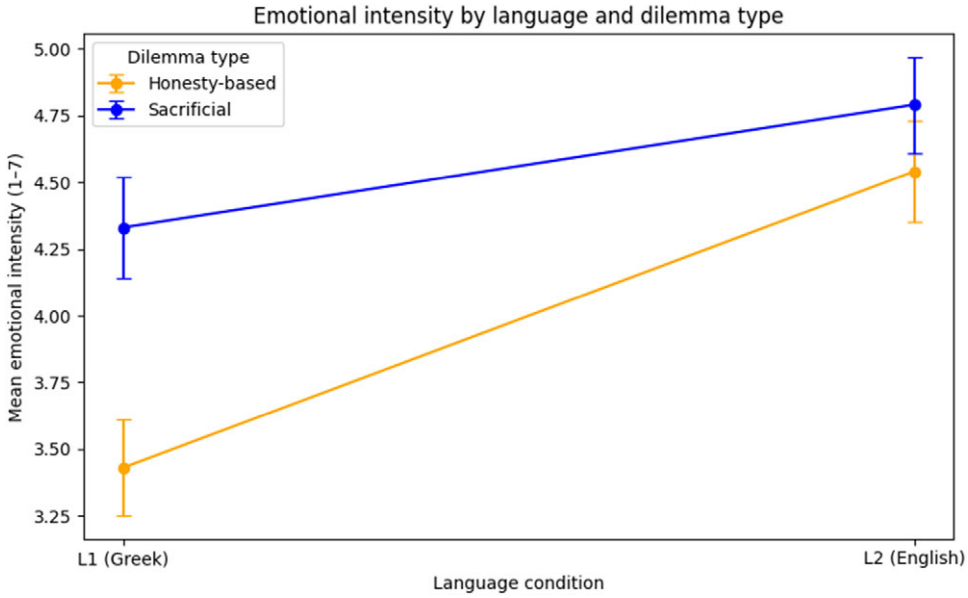


Figure 3. Emotional intensity as a function of language condition (L1 versus L2) and dilemma type (sacrificial-harm versus honesty-based interpersonal dilemmas). *Note:* Language condition is displayed as a categorical variable ('L1 (Greek)' and 'L2 (English)'). Separate lines represent sacrificial and honesty-based dilemma types. Points indicate group means, and error bars represent ±1 SE.

The results of the multilevel mediation analysis are presented in Table 5. The indirect effect of language on moral decisions through emotional intensity was small and did not reach statistical significance (ACME = −0.016, 95% CI [−0.039, 0.002], p = .070). Although emotional intensity was modelled as a predictor of moral decisions in the mediation analysis, it did not reliably mediate the effect of language on moral choices. In contrast, the direct effect of language on moral decisions reached statistical significance (ADE = 0.085, 95% CI [0.002, 0.165], p = .046). This finding suggests that language exerted a modest direct influence on action endorsement after accounting for emotional responses. However, the overall total effect of language on moral decisions was not statistically significant (p = .102), indicating that the overall magnitude of the language effect was small and should be interpreted with caution.

Taken together, these results indicate that emotional intensity and language both relate to moral decision-making, but emotional intensity does not appear to constitute the primary mechanism through which language influences moral choices in the present dataset.

Table 5. Multilevel mediation of language effects on moral decisions via emotion

Effect	Estimate	95% CI Lower	95% CI Upper	*p*-value
ACME	−0.016	−0.039	0.002	.070
ADE	0.085	0.002	0.165	.046*
Total effect	0.069	−0.013	0.147	.102
Prop. Mediated	−0.208	−1.403	1.229	.156

Note: *p < .05. ACME = Average Causal Mediation Effect. ADE = Average Direct Effect. Estimates are based on a multilevel mediation analysis with participant-level random intercepts in both the mediator and outcome models. Emotional intensity was modelled as the mediator, moral choice as the outcome, language condition as the treatment, and dilemma was included as a covariate.

7. Discussion

The current study investigated the role of moral identity in moral decision-making in a bilingual setting. To this end, Greek–English bilinguals completed the Moral Identity Questionnaire either in their L1 or in their L2, allowing us to examine whether language influenced their moral identity scores. We further examined whether these scores predicted participants' moral decisions and the intensity of their emotional responses to two sacrificial-harm and two honesty-based interpersonal moral dilemmas, as well as the mediating role of emotion in the relationship between language and moral choices.

Our results showed that language did not influence moral integrity but had a moderate to strong effect on moral self. Specifically, participants placed greater importance on being a moral person when completing the MIQ in their L1. This finding may reflect individuals' motivation to maintain close relationships and social harmony by being honest and fair and by avoiding harm to others; in other words, by embodying the characteristics of a righteous citizen. Indeed, individuals in collectivistic societies, such as Greece, tend to prioritise group welfare, view themselves as interdependent, and seek long-term relationships and strong emotional bonds with others (i.e., relatives or friends) by endorsing moral values such as honesty, trustworthiness, and conflict avoidance (Bhawuk, 2017; Graham et al., 2016). Furthermore, because education, the media, religion, and public moral discourse are typically experienced in a particular language (or languages), that language(s) may become closely linked to moral norms and personal memories. Given that the L1 is predominantly embedded in family life and is prevalent in sociomoral domains, it may therefore bring well-established moral values to the fore, thereby strengthening the salience of the moral self-concept. In contrast, languages that are less closely tied to these contexts may be associated with different moral representations, potentially resulting in attenuated moral self-evaluations (Peressotti et al., 2024).

It is also possible that the emotional resonance of certain L1 words in the Greek version of the MIQ (e.g., *ειλικρινής* 'honest' or

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

δίκαιος 'fair') may have led participants to report slightly higher scores than those on the equivalent items in the English version. For example, for an L1 Greek speaker, the phrase *είμαι ειλικρινής άνθρωπος* may be more emotionally charged than its English counterpart, *I am an honest person* (Pavlenko, 2005; Soriano & Ogarkova, 2025). Such vocabulary may have activated culturally embedded social norms associated with moral values. These norms are typically acquired during early childhood through real-life interactions that occur predominantly in L1 social contexts. Consequently, using the L1 is likely to elicit moral principles and heighten sensitivity to statements that contravene social and cultural norms to a greater extent than using an L2 (Gawinkowska et al., 2013). By contrast, language did not affect the extent to which participants valued acting in accordance with moral principles, that is, their moral integrity. This finding puts forward a new hypothesis: the *self* (in this study, the moral self) and how we project it to others may become more salient when we use our (more emotional) L1 than when we use an L2. Put differently, who we are – but not necessarily how we intend to act – may become less central in an L2 context.

Our findings also revealed that moral identity was a statistically significant predictor of our participants' emotions and moral decisions, with stronger effects observed in the honesty-based interpersonal dilemmas, regardless of the language used. Importantly, the main effects of moral identity were robust across three analytical levels: mixed-effects models using dilemma type, stimulus-level models treating each dilemma separately, and dilemma-specific regression analyses. One way to interpret this result is by considering prior research linking broader sociocultural influences to the development of moral identity. Although religiosity is only one such influence, studies have shown that individuals with some degree of religious affiliation tend to prefer deontological responses in emotionally charged situations (Saroglou & Craninx, 2021; Szekely et al., 2015). In Greece, religion retains a notable social presence through educational and community institutions (Halikioupoulou & Vasilopoulou, 2017), and survey data indicate high levels of religiosity and prosocial orientations (Papageorgiou, 2013; Pew Research Center, 2018). While these contextual factors help situate the cultural environment in which moral identity develops, they provide a theoretical backdrop rather than an empirical conclusion, as religiosity was not measured in the present study. The relative consistency of moral integrity across languages, together with the modest language effect on moral self, aligns with the idea that moral identity is shaped by enduring sociocultural conditions (Aquino & Reed, 2002; Hardy & Carlo, 2011). Nevertheless, future work incorporating explicit measures of religious commitment, cultural identification, or sociomoral value internalisation would be necessary to directly examine these underlying mechanisms.

The lack of a main effect of language on moral choices may lie in our participants' L2 experience. Factors such as L2 immersion (Mavrou et al., 2026; Winskel & Bhatt, 2020), L2 dominance (Privitera et al., 2023), and strong cultural associations with the L2 (Kyriakou & Mavrou, 2023b) appear to play a role in the magnitude of the MFLE. Although we did not empirically assess these variables, nor did we test the degree of acculturation of our participants, the intense exposure to and frequent use of English in the Greek education system, broadcast media, and cinema (Oikonomidis, 2003; Tatsioka, 2008) may have been sufficient to cancel out any direct language effects on moral decisions, as has been shown in previous studies conducted in Greece's neighbouring countries (see Kyriakou & Mavrou, 2023b).

With regard to emotion, participants in the L1 condition experienced stronger emotions in response to the sacrificial dilemmas than to the honesty-based dilemmas, whereas participants in the L2 condition provided significantly higher emotional ratings than their peers in the L1 condition in three out of four dilemmas (the vaccine, resume, and cheater's dilemmas). These results run counter to the hypothesis of increased emotional resonance in the L1 (as compared to the L2) but converge with recent findings by Del Maschio et al. (2026), in which participants also reported greater emotional distress when processing moral dilemmas and decision problems in their L2 than in their L1. The authors argued that the MFLE cannot be readily explained by emotional distancing and found limited support for an explanation based on increased cognitive effort. In particular, they suggested that the heightened feelings of distress, sadness, or worry observed in the L2 are difficult to reconcile with accounts based solely on cognitive load and therefore called for a reconsideration of the mechanisms underlying the MFLE. Our findings then appear to support the view that L2 users may overcompensate when providing emotion perception ratings by adjusting these ratings upwards. This phenomenon has already been observed in previous research (Dewaele & Moxsom-Turnbull, 2020; Lorette & Dewaele, 2024; Mavrou & Dewaele, 2020, among others), and the present study extends this finding to a moral decision-making context. These elevated emotional ratings in the L2 may partly reflect differences in how participants report or interpret emotional intensity across languages, not necessarily a direct increase in experienced emotion. Importantly, although emotional intensity was higher in the L2 condition, this increase did not translate into consistent or statistically reliable differences in moral decisions. Under this interpretation, these higher emotional ratings may not correspond to stronger affective engagement in a way that directly influences moral decisions.

The multilevel mediation analysis provided a more nuanced picture of the relationship between language, emotional intensity, and moral decisions. Although emotional intensity was modelled as a predictor of moral decisions, the estimated indirect effect of language through emotional intensity was small and not statistically significant. In contrast, a modest direct effect of language on moral decisions remained statistically significant after accounting for emotional responses, although the overall total effect of language on moral choices was not significant. Taken together, these results suggest that emotional intensity is linked to moral decision-making within the models but does not appear to be the primary mechanism through which language influences moral choices in the present data. Across analytical approaches, the effect of language on moral decisions was small and inconsistent and did not reliably vary as a function of dilemma type. This pattern suggests that differences in emotional intensity across language conditions may not translate directly into differences in behavioural responses. Moreover, differences in estimated effects across models likely reflect differences in model specification, particularly whether emotional intensity was included as a predictor, rather than evidence for different psychological processes.

This finding contributes to the ongoing debate regarding the mechanisms underlying the MFLE. While some accounts propose that reduced emotional resonance in a foreign language drives more utilitarian responses (e.g., Hayakawa et al., 2017; Kyriakou et al., 2023), other work has questioned whether emotional attenuation alone is sufficient to explain language-related differences in moral judgement (for reviews, see Circi et al., 2021; Del Maschio et al., 2022, 2026; Stankovic et al., 2022). The present findings align more closely with the latter view, suggesting that emotional intensity alone may not fully explain language-related differences in bilinguals' moral decisions.

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

An important conceptual issue concerns whether the sacrificial and honesty-based dilemmas used in this study capture the same underlying construct of moral decision-making. All four scenarios required participants to weigh competing moral norms and potential consequences and can therefore be considered moral dilemmas. However, the dilemmas differ in the type of moral conflict they evoke. The sacrificial-harm dilemmas follow the classic sacrificial format involving a direct clash between a deontological prohibition (e.g., not harming another person) and a utilitarian outcome, whereas the honesty-based dilemmas centre on honesty and interpersonal responsibility. The present design did not intend to isolate realism as a single psychological dimension. Rather, the contrast reflects differences between two moral contexts (sacrificial-harm dilemmas and honesty-based interpersonal dilemmas) and should therefore not be interpreted as a manipulation of realism *per se*. Additionally, the honesty-based dilemmas resemble those used in the emerging foreign-language literature on lying (e.g., Hu & Navarrete, 2025), suggesting that research on bilinguals' moral decision-making can extend beyond sacrificial trade-offs to include more routine social contexts. Thus, although both dilemma types elicit morally relevant reasoning, they likely engage partly distinct psychological processes: sacrificial-harm dilemmas involve high-conflict sacrificial reasoning, whereas honesty-based dilemmas involve compliance with everyday moral norms. In the latter case, judgements are more directly anchored in individuals' moral self-concepts and affective responses to interpersonal harm or deception, which may help explain why emotional intensity and moral identity were more strongly linked to moral decisions in the honesty-based dilemmas, despite the absence of a reliable mediation effect. Because each dilemma category in the present study was represented by only two scenarios, category-level interpretations should therefore be treated with caution. However, complementary analyses treating each dilemma as a separate fixed factor yielded a comparable overall pattern of results, suggesting that the observed effects were not driven by any single dilemma. Nevertheless, the present findings should be viewed as preliminary evidence requiring replication with larger and more diverse sets of moral stimuli.

Several additional limitations should also be acknowledged. The sample size was moderate, and we exclusively recruited Greek participants who conducted the different tasks either in their L1 Greek or in their L2 English. Including bilingual L1 English–L2 Greek speakers who completed the same tasks in both their L1 and L2 would provide richer data and allow for more robust within-participant comparisons. Although our preregistered power analysis indicated that a sample of 77 participants would be sufficient to detect medium-sized effects, the mixed-effects models used in the present study are more complex, incorporating random effects and interaction terms. As a result, the study is likely to have had limited power to detect small effects with respect to higher-order interactions. Additionally, the study relied on a small number of moral dilemmas (two sacrificial-harm and two honesty-based interpersonal scenarios). Since each dilemma category contained only two stimuli, the observed category-level patterns may partly reflect idiosyncratic features of the individual dilemmas. Although the exploratory nature of the study could justify this design, using only four dilemmas is unlikely to capture the full range of moral situations encountered in real life. Future research using larger and more diverse stimulus sets will be necessary to determine whether the patterns observed here generalise to a broader range of moral scenarios.

A further limitation related to the sample concerns the uneven gender distribution, with substantially fewer male than female participants. Gender was not included as a control variable in the main analyses for two reasons: first, the small number of male participants (n = 20) would have limited the stability and interpretability of any gender-related effects; and second, the central focus of the study was on moral identity and language, and gender was not part of our theoretical predictions. Nonetheless, future research with more balanced samples would be valuable in examining whether gender contributes to variability in moral identity, emotional responses, or moral decisions across languages. Another limitation concerns the age imbalance between the two language groups. Although the L1 and L2 groups differed in age, supplementary analyses demonstrated that age was not associated with emotional intensity or moral choices, and the main findings were replicated within a more homogeneous subsample (participants aged 18–39). Thus, age is unlikely to have influenced the results, although future studies would benefit from more balanced age distributions.

Additionally, although the MIQ was carefully translated, cross-language differences in self-report measures may reflect not only differences in the underlying construct but also differences in how participants interpret, access, and express self-relevant information in different languages. This is especially important for moral identity, which involves abstract and self-referential concepts that may vary in emotional resonance, accessibility, or personal salience across L1 and L2. Consequently, the lower moral self scores observed in the L2 condition should be interpreted with caution. While this finding is consistent with the possibility that language context shapes the accessibility of moral aspects of the self, alternative measurement-related explanations cannot be completely ruled out. However, only the moral self dimension differed across language conditions, whereas moral integrity remained stable. This pattern may suggest that some components of moral identity are more sensitive than others to language context. Future studies thus would benefit from formally testing measurement invariance across language versions of the MIQ before interpreting mean-level differences as evidence of language-related changes in moral identity. Combining self-report measures with behavioural, implicit, or narrative indicators of moral identity would also help clarify whether language affects moral identity or its self-reported expression.

A further consideration is that emotion ratings were always presented before the moral decision question. This fixed order was necessary to maintain the temporal sequence required for the mediation analysis, but it may also have increased participants' attention to their emotional reactions. Future studies could use designs that balance the need for temporal ordering with appropriate counterbalancing procedures. A related limitation is that emotion and moral decisions were treated as separate variables, but we cannot rule out the possibility that emotion is 'part of' the overall moral judgement. To address this possibility, future studies could use composite scores reflecting both emotion and decision. This is in line with Hardy and Carlo's (2005) observation that cognition, emotion, and identity are overlapping constructs, highlighting the need to use diverse and creative methodologies to examine how they interact with each other.

Another methodological consideration concerns the format of the emotional rating scale. We used 7-point Likert scales with verbal endpoint labels, which can sometimes elicit the Anchor Contraction Effect (De Langhe et al., 2011), whereby respondents shift ratings toward the middle of the scale, particularly when linguistic processing of the anchors is more effortful in an L2. This phenomenon may provide an alternative explanation for the

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

elevated emotional ratings observed in the L2 condition. Although prior work has also documented over-reporting of emotional intensity in an L2, future studies would benefit from using non-verbal or continuous rating formats (e.g., sliders or pictorial scales) to minimise anchoring effects and better isolate language-related differences in emotionality. This measurement-related explanation may also help explain the apparent dissociation between emotional intensity and moral decisions in the present study. Although participants reported higher emotional intensity in the L2 condition, these elevated ratings did not correspond to systematic differences in moral choices. This suggests that differences in emotional ratings across languages may not necessarily reflect differences in affective processes that directly influence decision-making but may instead arise, at least partly, from variation in how emotional experiences are interpreted and reported.

The mediation analysis employed in the current study is also open to potential criticism. This kind of analysis assumes that the variables involved are measured without error – a challenging requirement when dealing with subjective emotional states – and should be ideally supported by longitudinal designs (Pek & Hoyle, 2016). Although the mediation model in the present study offers useful insights, future studies should aim to employ measurement approaches that minimise error and strengthen the robustness of emotion-related constructs. As emotional intensity was operationalised using self-reported ratings, potential measurement-related biases in these ratings may have affected the mediation estimates. This possibility should thus be considered when interpreting the indirect and direct effects.

Moreover, the distinction between deontology and utilitarianism may not always map neatly onto all types of moral dilemmas. In the case of the two honesty-based dilemmas used in the present study, and without intending to generalise beyond them, behaviours such as cheating on someone or lying may be regarded as ethically incorrect.[11] Recent studies have further revealed that moral dilemmas involving cheating or lying can elicit heightened levels of negative emotions, such as guilt and regret (Kyriakou et al., 2024; Kyriakou & Mavrou, 2023a). However, there are multiple possibilities depending on the approach that each study or researcher adopts. From a utilitarian perspective, for example, lying to achieve a desirable outcome (e.g., obtaining a better job) or to protect a loved one from psychological harm may be seen as the right thing to do. That is, lying is morally acceptable when the resulting consequences maximise benefit (i.e., getting hired) or minimise harm (i.e., negative feelings, depression, and the disappointment of not getting the job). However, similarly to deontologists, utilitarians could also say that lying is bad or immoral because of the worse consequences it has compared to telling the truth and because it may harm one's character (see Carson, 2005, for a discussion). It becomes clear, then, that decision-making is not simply a matter of black and white. Future research should therefore consider more thoroughly the reasons and arguments underlying moral reasoning, using qualitative approaches such as those adopted by Kyriakou et al. (2023).

8. Conclusion

The current study provides novel evidence that moral identity influences emotional responses and tendencies to reject or endorse actions, particularly when bilinguals confront everyday honesty-based moral dilemmas. At the same time, the influence of language on moral decisions appears more nuanced than commonly assumed in the MFLE literature. While the use of the L1 heightened the importance of the moral self, the use of the L2 neither diminished moral integrity nor consistently promoted greater endorsement of outcome-based actions in sacrificial dilemmas or greater action endorsement in honesty-based dilemmas. Instead, language-related effects emerged primarily in emotional intensity, with L2 use associated with amplified emotional ratings. This pattern aligns with recent work suggesting that self-reported emotions in an L2 may be inflated rather than muted, offering important methodological and theoretical implications for interpreting bilinguals' emotional data in moral contexts. The mediation analysis further indicated that emotional intensity did not reliably mediate the relationship between language and moral judgement, although a small direct effect of language on moral decisions was observed after accounting for emotional responses. Taken together, these findings underscore the importance of integrating identity-level factors and language-related processes into models of bilingual moral cognition. Moral decision-making in multilingual adults is unlikely to be fully captured by a simple contrast between emotionality in the L1 and detachment in the L2. Rather, it reflects how culturally grounded yet individually instantiated moral identities interact with linguistic experience and the specific demands of different moral dilemmas. Recognising this complexity is crucial for advancing more comprehensive theories of bilingual decision-making.

These insights also carry relevance for applied settings. Understanding the ways in which young bi−/multilinguals reason and decide across their languages may inform broader discussions about the development of moral reflection and critical thinking in multilingual contexts. While our findings do not speak directly to the design of educational interventions, they do suggest that moral decision-making is shaped by both relatively stable identity constructs and language-related processing differences. These influences may therefore be relevant to future work exploring how moral reasoning can be supported within educational settings. In a globalised world where moral decisions are often made in *lingua francas* such as English, recognising the subtle shifts introduced by moral identity and language is equally important for understanding how individuals navigate professional, academic, and socio-political contexts.

Data availability statement. The preregistration, study materials, raw data, supplementary files, and analysis syntax supporting the findings of this study are openly available on the Open Science Framework (OSF) at: https://osf.io/emvaz/.

Competing interests. The authors report that there are no competing interests to declare.

[11]We do not assume that lying is inherently wrong (see Carson, 2005, for a discussion); rather, we interpret our findings in relation to the specific dilemmas used in the study.

References

Aquino, K., Freeman, D., Reed, A. I. I., Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, **97**(1), 123–141. https://doi.org/10.1037/a0015406

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, **83**(6), 1423–1440. http://doi.org/10.1037//0022-3514.83.6.1423

Aquino, K., Reed, A. I. I., Thau, S., & Freeman, D. (2007). A grotesque and dark beauty: How moral identity and mechanisms of moral disengagement influence cognitive and emotional reactions to war. *Journal of Experimental Social Psychology*, **43**(3), 385–392. https://doi.org/10.1016/j.jesp.2006.05.013

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

Bates, D., **Mächler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauman, C. W., **McGraw, A. P.**, **Bartels, D. M.**, & **Warren, C.** (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, **8**(9), 536–554. https://doi.org/10.1111/spc3.12131

Bhawuk, D. P. S. (2017). Individualism and collectivism. In Y. Y.Kim & K. L. McKay-Semmler (Eds.), *The international encyclopedia of intercultural communication* (Vol. **2**, pp. 920–929). Wiley. https://doi.org/10.1002/9781118783665.ieicc0107

Białek, M., **Paruzel-Czachura, M.**, & **Gawronski, B.** (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, **85**, 103855. https://doi.org/10.1016/j.jesp.2019.103855

Black, J. E., & **Reynolds, W. M.** (2016). Development, reliability, and validity of the moral identity questionnaire. *Personality and Individual Differences*, **97**, 120–129. https://doi.org/10.1016/j.paid.2016.03.041

Blasi, A. (1983). Moral cognition and moral action: A theoretical perspective. *Developmental Review*, **3**(2), 178–210. https://doi.org/10.1016/0273-2297(83)90029-1

Caldwell-Harris, C. L. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science*, **24**(3), 214–219. https://doi.org/10.1177/0963721414566268

Carson, T. L. (2005). Ross and utilitarianism on promise keeping and lying: Self-evidence and the data of ethics. *Philosophical Issues*, **15**(1), 140–157. https://doi.org/10.1111/j.1533-6077.2005.00058.x

Christensen, J. F., **Flexas, A.**, **Calabrese, M.**, **Gut, N. K.**, & **Gomila, A.** (2014). Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology*, **5**, 607. https://doi.org/10.3389/fpsyg.2014.00607

Christensen, R. H. B. (2023). *ordinal: Regression models for ordinal data* (R package version 2023.12–4) [Computer software]. https://CRAN.R-project.org/package=ordinal

Circi, R., **Gatti, D.**, **Russo, V.**, & **Vecchi, T.** (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, **28**(4), 1131–1141. https://doi.org/10.3758/s13423-020-01871-z

Colby, A., & **Damon, W.** (1992). *Some do care: Contemporary lives of moral commitment*. Free Press.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, **78**(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

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. https://doi.org/10.1371/journal.pone.0094842

De Langhe, B., **Puntoni, S.**, **Fernandes, D.**, & **van Osselaer, S. M. J.** (2011). The anchor contraction effect in international marketing research. *Journal of Marketing Research*, **48**(2), 366–380. https://doi.org/10.1509/jmkr.48.2.366

Del Maschio, N., **Bellini, C.**, **Abutalebi, J.**, & **Sulpizio, S.** (2026). Boundaries of the foreign language effect under conditions of uncertainty. *Bilingualism: Language and Cognition*, **29**(3), 784–794. https://doi.org/10.1017/S1366728925100400

Del Maschio, N., **Crespi, F.**, **Peressotti, F.**, **Abutalebi, J.**, & **Sulpizio, S.** (2022). Decision-making depends on language: A meta-analysis of the foreign language effect. *Bilingualism: Language and Cognition*, **25**(4), 617–630. https://doi.org/10.1017/S1366728921001012

Dewaele, J.-M., **Mavrou, I.**, **Kyriakou, A.**, & **Lorette, P.** (2024). The role of language and emotional intelligence in judgments of real-life social and moral transgressions among Greek, Hungarian, and British users of English. *Journal of Behavioral Decision Making*, **37**(1), e2356. https://doi.org/10.1002/bdm.2356

Dewaele, J.-M., & **Moxsom-Turnbull, P.** (2020). Visual cues and perception of emotional intensity among L1 and LX users of English. *International Journal of Multilingualism*, **17**(4), 499–515. https://doi.org/10.1080/14790718.2019.1612902

Dylman, A. S., & **Champoux-Larsson, M.-F.** (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

EF Education First. (2022). *EF English Proficiency Index (EF EPI)*. https://www.ef.com/epi/

European Commission. (2025). *Foreign language skills statistics*. Eurostat.

Faul, F., **Erdfelder, E.**, **Buchner, A.**, & **Lang, A.-G.** (2009). *Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. Behavior Research Methods*, **41**(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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**(2), 175–191. https://doi.org/10.3758/BF03193146

Gawinkowska, M., **Paradowski, M. B.**, & **Bilewicz, M.** (2013). Second language as an exemptor from sociocultural norms. Emotion-related language choice revisited. *PLoS One*, **8**(12), e81225. https://doi.org/10.1371/journal.pone.0081225

Geipel, J., **Hadjichristidis, C.**, & **Surian, L.** (2015a). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, **59**, 8–17. https://doi.org/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. https://doi.org/10.1371/journal.pone.0131529

Goh, T.-T., **Dai, X.**, **Liu, I. Q.**, & **Wen, B.** (2022). Heightening of moral identity: The role of mobile prosocial messages and presence. *Journal of Moral Education*, **51**(2), 139–154. https://doi.org/10.1080/03057240.2020.1793743

Graham, J., **Meindl, P.**, **Beall, E.**, **Johnson, K. M.**, & **Zhang, L.** (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, **8**, 125–130. https://doi.org/10.1016/j.copsyc.2015.09.007

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. https://doi.org/10.1126/science.1062872

Grover, S. L. (1993)**.** Lying, deceit, and subterfuge: A model of dishonesty in the workplace. *Organization Science*, **4**(3), 478–495. https://doi.org/10.1287/orsc.4.3.478

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, **108**(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Halikioupoulou, D., & **Vasilopoulou, S.** (2017). Political instability and the persistence of religion in Greece: The policy implications of the cultural defence paradigm. In F. C. González & G. D'Amato (Eds.), *Multireligious society: Dealing with religious diversity in theory and practice* (pp. 63–76). Routledge.

Hambleton, R. K., **Merenda, P. F.**, & **Spielberger, C. D.** (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Lawrence Erlbaum Associates.

Hardy, S. A., & **Carlo, G.** (2005). Identity as a source of moral motivation. *Human Development*, **48**(4), 232–256. https://doi.org/10.1159/000086859

Hardy, S. A., & **Carlo, G.** (2011). Moral identity: What is it, how does it develop, and is it linked to moral action? *Child Development Perspectives*, **5**(3), 212–218. https://doi.org/10.1111/j.1750-8606.2011.00189.x

Hayakawa, S., **Tannenbaum, D.**, **Costa, A.**, **Corey, J. D.**, & **Keysar, B.** (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, **28**(10), 1387–1397. https://doi.org/10.1177/0956797617720944

Hu, Z., & **Navarrete, E.** (2025). Understanding the impact of foreign language on social norms through lies. *Bilingualism: Language and Cognition*, **28**(1), 257–271. https://doi.org/10.1017/S1366728924000373

JASP Team. (2024). *JASP* (Version 0.95.0) [Computer software]. https://jasp-stats.org/

Kahane, G., **Everett, J. A. C.**, **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. https://doi.org/10.1016/j.cognition.2014.10.005

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., **Hayakawa, S. L.**, & **An, S. G.** (2012). The foreign-language effect. *Psychological Science*, **23**(6), 661–668. https://doi.org/10.1177/0956797611432178

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

Koenigs, M., **Young, L.**, **Adolphs, R.**, **Tranel, D.**, **Cushman, F.**, **Hauser, M.**, & **Damasio, A.** (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, **446**(7138), 908–911. https://doi.org/10.1038/nature05631

Kyriakou, A., **Foucart, A.**, & **Mavrou, I.** (2023). Moral judgements in a foreign language: Expressing emotions and justifying decisions. *International Journal of Bilingualism*, **27**(6), 978–995. https://doi.org/10.1177/13670069221134193

Kyriakou, A., & **Mavrou, I.** (2023a). ¿Eres muy emocional? I don't think so. How does language influence our emotional responses to everyday moral dilemmas? In A. B.Canales & S. M. Leralta (Eds.), *Emotion and identity in foreign language learning* (pp. 297–321). Peter Lang.

Kyriakou, A., & **Mavrou, I.** (2023b). What language does your heart speak? The influence of foreign language on moral judgements and emotions related to unrealistic and realistic moral dilemmas. *Cognition and Emotion*, **37**(8), 1330–1348. https://doi.org/10.1080/02699931.2023.2258577

Kyriakou, A., & **Mavrou, I.** (2025). I can't kill them, but I can throw them over the bridge: Does the emotionality of moral questions influence bilinguals' moral judgements? *Bilingualism: Language and Cognition*, **28**(4), 1044–1055. https://doi.org/10.1017/S1366728924000877

Kyriakou, A., **Mavrou, I.**, & **Palapanidi, K.** (2024). The role of foreign language in the experience and emotional expression of guilt: Evidence from moral scenarios and autobiographical memories of bilinguals. *International Journal of Bilingual Education and Bilingualism*, **27**(10), 1407–1421. https://doi.org/10.1080/13670050.2024.2365238

Lee, M. D., & **Wagenmakers, E.-J.** (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lemhöfer, K., & **Broersma, M.** (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lorette, P., & **Dewaele, J. M.** (2024). Interpersonal perception of emotional intensity by English first (L1) and foreign (LX) language users in audio(visual) communication. *International Journal of Multilingualism*, **21**(2), 1116–1133. https://doi.org/10.1080/14790718.2022.2144326

Mavrou, I., & **Dewaele, J. M.** (2020). Emotionality and pleasantness of mixed-emotion stimuli: The role of language, modality, and emotional intelligence. *International Journal of Applied Linguistics*, **30**(2), 313–328. https://doi.org/10.1111/ijal.12285

Mavrou, I., **Revesz, A.**, **Kyriakou, A.**, & **Rong, X.** (2026). Cognition and emotion in moral decision-making: The role of working memory, emotional intelligence, and language factors in bilinguals' moral judgements and emotions. *Studies in Second Language Acquisition*, **48**(2), 424–449. https://doi.org/10.1017/S0272263125101484

McNab, N., & **Mavrou, I.** (2025a). Emotion and moral stance in evaluations of impoliteness in L1 and L2 from video clips of workplace interactions. *Lingua*, **322**, 103949. https://doi.org/10.1016/j.lingua.2025.103949

McNab, N., & **Mavrou, I.** (2025b). How we are versus how we are feeling: The role of emotional intelligence and mood in reactions to impoliteness in L1 and L2. *Journal of Pragmatics*, **246**, 121–133. https://doi.org/10.1016/j.pragma.2025.07.001

Miozzo, M., **Navarrete, E.**, **Ongis, M.**, **Mello, E.**, **Girotto, V.**, & **Peressotti, F.** (2020). Foreign language effect in decision-making: How foreign is it? *Cognition*, **199**, 104245. https://doi.org/10.1016/j.cognition.2020.104245

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, **15**(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Oikonomidis, A. (2003). The impact of English in Greece. *English Today*, **19**(2), 55–61. https://doi.org/10.1017/S0266078403002104

Papageorgiou, N. (2013). Belonging without believing?" dimensions of the religious phenomenon in modern Greece. *Synthesis*, **2**(1), 176–186. https://doi.org/10.26262/syn.v2i1.3661

Pavlenko, A. (2005). *Emotions and multilingualism*. Cambridge University Press.

Pek, J., & **Hoyle, R. H.** (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass*, **10**(3), 150–163. https://doi.org/10.1111/spc3.12237

Peressotti, F., **Lorenzoni, A.**, & **Miozzo, M.** (2024). Moral judgments in native, regional, and foreign languages. *Bilingualism: Language and Cognition*, **27**(1), 154–163. https://doi.org/10.1017/S1366728923000342

Pew Research Center. (2018). *Eastern and Western Europeans differ on importance of religion, views of minorities, and key social issues.* Pew Research Center. https://www.pewresearch.org/religion/2018/10/29/eastern-and-western-europeans-differ-on-importance-of-religion-views-of-minorities-and-key-social-issues/

Posit Team (2022). *RStudio: Integrated Development Environment for* [Computer software]. Posit Software, PBC. URL http://www.posit.co/

Privitera, A. J., **Li, S.**, **Zhou, Y.**, & **Wang, M.** (2023). Modulatory role of foreign language experience on the moral foreign language effect. *Bilingualism: Language and Cognition*, **26**(5), 1038–1050. https://doi.org/10.1017/S1366728923000275

Qualtrics. (2022). *Qualtrics survey software* [Computer software]. Qualtrics International Inc. https://www.qualtrics.com

Reynolds, S. J., & **Ceranic, T. L.** (2007). The effects of moral judgment and moral identity on moral behavior: An empirical examination of the moral individual. *Journal of Applied Psychology*, **92**(6), 1610–1624. https://doi.org/10.1037/0021-9010.92.6.1610

Rhemtulla, M., **Brosseau-Liard, P. É.**, & **Savalei, V.** (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, **17**(3), 354–373. https://doi.org/10.1037/a0029315

Rouder, J. N., & **Morey, R. D.** (2012). Default bayes factors for model selection in regression. *Multivariate Behavioral Research*, **47**(6), 877–903. https://doi.org/10.1080/00273171.2012.734737

Rouder, J. N., **Speckman, P. L.**, **Sun, D.**, **Morey, R. D.**, & **Iverson, G.** (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, **16**(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sakellariou, A. (2013). Religion in Greek society state, public or private? In W. Hofstee & A. van der Kooij (Eds.), *Religion beyond its private role in modern society* (pp. 153–166). Brill. https://doi.org/10.1163/9789004257856_010

Saroglou, V., & **Craninx, M.** (2021). Religious moral righteousness over care: A review and a meta-analysis. *Current Opinion in Psychology*, **40**, 79–85. https://doi.org/10.1016/j.copsyc.2020.09.002

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, **8**(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Sommer, M., **Rothmayr, C.**, **Döhnel, K.**, **Meinhardt, J.**, **Schwerdtner, J.**, **Sodian, B.**, & **Hajak, G.** (2010). How should I decide? The neural correlates of everyday moral reasoning. *Neuropsychologia*, **48**(7), 2018–2026. https://doi.org/10.1016/j.neuropsychologia.2010.03.023

Soriano, C., & **Ogarkova, A.** (2025). The meaning of 'frustration' across languages. *Language and Cognition*, **17**, e16. https://doi.org/10.1017/langcog.2024.50

Stankovic, M., **Biedermann, B.**, & **Hamamura, T.** (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

Stets, J. E. (2010). The social psychology of the moral identity. In S. Hitlin & S. Vaisey (Eds.), *Handbook of the sociology of morality* (pp. 385–409). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-6896-8_20

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, **80**(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Szekely, R. D., **Opre, A.**, & **Miu, A. C.** (2015). Religiosity enhances emotion and deontological choice in moral dilemmas. *Personality and Individual Differences*, **79**, 104–109. https://doi.org/10.1016/j.paid.2015.01.036

Tatsioka, Z. (2008). Code-switching and English loanwords in the Greek media. *Newcastle Working Papers in Linguistics*, **14**, 129–144.

Tingley, D., **Yamamoto, T.**, **Hirose, K.**, **Keele, L.**, & **Imai, K.** (2013). *mediation: R package for causal mediation analysis* (R package version 4.4.2) [Computer software]. http://CRAN.R-project.org/package=mediation

Tingley, D., **Yamamoto, T.**, **Hirose, K.**, **Keele, L.**, & **Imai, K.** (2014). Mediation: R Package for causal mediation analysis. *Journal of Statistical Software*, **59**(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Van de Vijver, F. J. R., & **Hambleton, R. K.** (1996). Translating tests: Some practical guidelines. *European Psychologist*, **1**(2), 89–99. https://doi.org/10.1027/1016-9040.1.2.89

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press

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. https://doi.org/10.1371/journal.pone.0203528

Warner, C. H., **Fortin, M.**, & **Melkonian, T.** (2024). When are we more ethical? A review and categorization of the factors influencing dual-process ethical decision-making. *Journal of Business Ethics*, **189**(4), 843–882. https://doi.org/10.1007/s10551-022-05281-0

Winskel, H., & **Bhatt, D.** (2020). The role of culture and language in moral decision-making. *Culture and Brain*, **8**(2), 207–225. https://doi.org/10.1007/s40167-019-00085-y

Winterich, K. P., **Mittal, V.**, & **Aquino, K.** (2015). Moral and political identity. In M. I. Norton, D. D. Rucker, & C. Lamberton (Eds.), *The Cambridge handbook of consumer psychology* (pp. 589–618). Cambridge University Press. https://doi.org/10.1017/CBO9781107706552.022

https://doi.org/10.1017/S136672892610159X Published online by Cambridge University Press