The Foreign Language Effect (FLE) on moral judgment and the role of foreign language proficiency and emotionality.

This thesis is submitted to the Department of Linguistics and English Language of Lancaster University for the Degree of PhD Linguistics

By

Nikki-Maria Christofi

Thesis Supervised by Dr. Aina Casaponsa Galí

Lancaster University- UK, September 2024

To my parents

... for always being there for me.

The Foreign Language Effect (FLE) on moral judgment and the role of foreign language proficiency and emotionality.

By Nikki-Maria Christofi

Abstract

This study aims to broaden the understanding of the phenomenon of the foreign language effect (FLE; the systematic influence of a foreign language on decision making; Dylman, & Champoux-Larsson 2020) on our moral decision and judgements. This rising literature topic focuses on how the language we speak when making choices can affect our decisions and moral judgment, however, the precise reasons that the FLE occurs are unclear (Hayakawa et al., 2016). Is it due to disfluency that requires more cognitive effort in processing information in a foreign language or is it due to the reduced emotionality due to the nature in which a second language (L2) is learnt? To provide a clearer image on the role of cognitive effort and the role of reduced emotionality on the FLE the current study was designed in a way that will weight evidence on each scope.

Since the cognitive load hypothesis is based on the assumption of disfluency and poorer L2 proficiency, the study addressed this by introducing rigorous proficiency measures, as well as standard fluency tests in the first experiment . The second experiment explores in more depth the role of emotionality, specifically on emotionally charged moral decisions and judgements. Hence, to test of the role of reduced emotionality on the FLE (Hayakawa et al., 2016) more rigorously, the current study employed a validated emotion measure PANAS-X; Positive and Negative Affect Schedule, (Horne et. al, 2016) prior to (Pre-test Emotion Measure) and following (Post-Test Emotion Measure) moral judgements.

Conclusive, when measuring proficiency rigorously it does not seem to be a factor that moderates the FLE, however what seems to be driving the FLE phenomenon is language emotionality, specifically, negative affect and emotions of hostility are more prominent in the L1 rather than the L2 after participants have been exposed to the dilemmas.

Contents

List of Figures

Tables

Acknowledgments

This PhD thesis has been one of the most important challenges and one of my biggest and most exciting goals in life. It has been a demanding journey combined with working full-time and with life challenging situations that taught me that you need to keep going and never give up no matter what life throws at you. Through these words I hope that I'll be able to express my gratitute to the people who made the completion of this thesis possible.

Firstly, I would like to thank my supervisors for their invaluable guidance and support during the course of my PhD journey at Lancaster University. I want to express my sincere appreciation to my first and former supervisor, Professor Panos Athanasopoulos, for his unwavering belief in me from the very beginning, along with his constant guidance and encouragement for as long as he served as my supervisor. When I decided to pursue a PhD in Linguistics and contacted him with the prospect of being my PhD supervisor, he had no hesitation in recommending my application for approval and supervise the project, having known my work as my master's thesis supervisor in the past. Thank you for your overall invaluable support and your presence at my first oral presentation at the international EuroSLA32 conference, where I presented my first major experiment. Your remarkable mentorship has been a cornerstone of my academic journey. Ευχαριστώ από καρδιάς!

I am equally grateful to my current supervisor Dr. Aina Casaponsa, who Panos suggested in co-supervising this PhD around the second year, as her expertise in the field of Psycholinguistics offered very helpful additional insights to this research. Her exceptional guidance and thorough feedback helped me design and complete the second experiment of this study and provided me with useful advice on all aspects of the PhD. Despite her demanding schedule she instantly agreed to be my sole supervisor, after Panos relocated from the UK, and provided me with important support, which was crucial for the successful completion of this thesis. Muchas Gracias Aina!

I would also like to thank my appraisal panel members; Professor Patrick Rebuschat for acting as my confirmation panelist, offering very useful advice and

motivation throughout his report, and Silke Brandt for acting as my pre and post confirmation panel member offering an extra confirmation on the progress of my work.

Foremost, I would like to express my deepest gratitude to my family. To my parents, Chris and Christina Christofi, for the unwavering support, unconditional love, and their constant presence that has been my greatest source of strength throughout these years. A special thank you to my father for his important assistance in recruiting participants for the data collection in all experiments and most of all for motivating me to embark on this journey, supporting me and always believing in me. I would also like to thank my sisters, Dora and Demetra for always being there for me. Demetra with her expertise as an English language teacher helped me proofread parts of the PhD and reviewed the translation of the research instruments from English to Greek along with Marina.

My deepest thanks to my dearest friend Marina Ioannou, who has always been there for me during life challenges and her constant motivation to never give up. I am also deeply grateful for her help with pilot testing and assistance in recruiting participants. I am also particularly grateful to my colleagues Eraklis Aristidis for his support and advice and to Amvrosios Prodromou for his encouragement to begin this PhD. I would also like to express my heartfelt appreciation to Stephanos Panteli for his unwavering support during the final stages of my PhD. I would also like to thank Nikolas Frangos for his IT support when needed. I extend my gratitude to all the friends who have stood by me in any way throughout these years. Those who offered a listening ear when I enthusiastically talked about my research and findings sharing their interest and insights. I am truly thankful for your support.

I also wish to express my sincere appreciation to all the participants who took part in this study, particularly those involved in experiment 1, who devoted significant time in completing the lengthy questionnaires.

Finally, I acknowledge the Peel Trust for awarding me the studentship fund for this academic year and Lancaster University for offering me this incredible opportunity to pursue my academic aspirations.

Author's declaration

I hereby declare that the work presented in this thesis is my own and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

I would like to acknowledge that experiment 1 was presented in an oral presentation at the conference of the European Second Language Association (EuroSLA 32) that was held in August 2023 (Christofi et al., 2023) and the abstract presented was identical to the one presented at the conference.

Furthermore, all experiments of this thesis have been written in the form of articles submitted to respected journals for review. Experiment 2 was also made available at PsyArXiv Preprints.

Chapter 1. Introduction

A foreign language (FL) is one that individuals acquire and utilize in settings where most speakers do not have it as their native tongue, and it's employed in contexts where it's not the primary language for the majority of speakers (Pavlenko, 2012). Contemporary globalization has established circumstances where individuals proficient in multiple languages might be anticipated to make choices using a language other than their native one. Surprisingly, these bilingual individuals do not consistently make the same decisions across languages; instead, their choices vary depending on the language used, a phenomenon termed the Foreign Language Effect; FLE (Privitera et al., 2023). This is one of the most cutting-edge topics in the field of bilingualism, the so-called FLE: Does thinking in our second (L2)/FL make us more efficient thinkers, leading to more logical, utility-driven decisions? (see McFarlane et al., 2020; e.g., Hayakawa et al., 2016).

One of the first studies to demonstrate variations in decision-making outputs based on the language bilingual individuals used focused on moral decision making. In moral decision making it is often observed that some individuals are more inclined to making utilitarian choices when doing these tasks in their L2 rather than in their L1. This means prioritizing outcomes that maximize overall benefit, even if it requires difficult moral decision making. Specifically, Keysar, Hayakawa, and An (2012) were the first to provide empirical evidence that bilinguals were more likely to make utilitarian decisions when operating in their L2 compared to their L1. This study was among the first to establish that the language of operation can influence decision making. The FLE has also been found in other domains of decision making, such as risk-taking (Dylman & Champoux-Larsson, 2020; Hadjichristidis et al., 2015; Hadjichristidis, et al.,2019) and emotional decisions (Dylman & Bjärtå 2019; García-Palacios et al., 2018; Vives et al.,2021; Zheng et al.,2020) and was further examined by Costa et al. (2014), who explored the interplay between language and cognition.

In this thesis, I will delve into whether processing written information in a foreign language, as opposed to one's native tongue, influences moral judgment and decision-making in the sphere of moral reasoning. Moral reasoning refers to the process by which individuals make decisions about what is right and wrong, by evaluating actions, motives, and principles based on ethical standards (Smith, 2020). In the context of moral judgment, the term refers to the evaluation of actions and character of others. The examination of circumstances under which individuals choose to act morally or immorally, applying consistent standards to themselves and others, constitutes a rich and intriguing area of research. Generally, moral judgments pertain to actions where one party either harms or aids another or treats individuals or groups with fairness or unfairness. Additionally, moral judgments may encompass behaviours deemed morally relevant by certain individuals but not by others (Avramova & Inbar, 2013).

A widely used approach on measuring moral judgement in previous research that explores the FLE is moral dilemmas (hypothetical moral dilemmas that require participants to hypothetically kill one person to save 5-10 other individuals from certain death). In this regard, previous studies have shown that making decisions in the FL leads to more utilitarian responses, kill one person to save the many (e.g. Wong & Ng., 2018). Yet, one of the great unknowns in FLE research is whether the effect is rooted in the emotional reduction that occurs when an FL is used; emotion-reducing hypothesis (**potential Factor 1)**, or from more processing effort resulting from the lower proficiency in the FL known as the cognitive load hypothesis (**potential Factor 2).**

Potential Factor 1: Emotional reduction in the FL

The first possible factor that it is argued to drive the FLE regards to the emotional reduction. The reduced emotionality account posits that our emotions wield a more pronounced sway over our decision-making processes when operating in our native language compared to a non-native one (Keysar et al., 2012; Corey et al., 2017; Hayakawa et al., 2017; Vives et al., 2018; Hadjichristidis et al., 2019). This theory suggests that since emotions can potentially influence certain cognitive shortcuts or biased reasoning, deliberately thinking in a non-native language may diminish emotional interference, thus allowing for more uninterrupted and undistorted reasoning

compared to when reasoning in one's native language (McFarlane et al., 2020). Certain theories could be directly linked to this account. For example, based on the dual-process theory there are two types of processing; the first type is rapid instant and non-conscious, but the second type takes longer and is systematic (Cipoletti et al., 2016). The first type is more likely to involve decision making in the L1 (where emotions are involved) and the second type is more likely to involve decision making in an FL (more neutral less affected by emotions; see Keysar et al., 2012; Cushman, 2013).

This factor will be explored in experiment 2 of this study that will focus solely on the role of the language emotionality on moral judgement on the FLE by exploring specific positive and negative emotions (e.g. Cordellieri et al., 2020) that could be induced in experiments that require moral decision making on the FLE. Language emotionality on the FLE has mostly been elicited using one question that asked participants to evaluate how distressed they feel after making a moral decision on a hypothetical dilemma (Wong & Ng., 2018). Therefore, in experiment 2 of the current PhD I will employ systematic emotion rating measures in order to explore the role of different emotional categories on the FLE.

It is stressed that apart from the emotional reduction in the FL there are other factors that "mediate or moderate" (p.3) the FLE on moral decision-making and stress that future research should thoroughly examine participants' background factors that may influence the reduced emotional effect, such as how proficient a participant is in the FL (Hayakawa et al.2016) which leads to potential factor 2 as follows.

Potential Factor 2: Cognitive load hypothesis

According to Hayakawa et al., (2017) disfluency in an FL could be an important factor of FLE, as processing disfluency could lead to, a more deliberative mode of thinking, as the level of difficulty could result in more careful processing. Nonetheless, at the same time FL proficiency has been poorly measured (see studies in Circi et al., 2021 meta-analysis), and verbal fluency has not been measured at all. FL proficiency, while hypothesized to underlie the processing effort account (e.g., Hayakawa et al.,

2017), has so far not been rigorously measured. This leads to the novelty of the first study (experiment 1) where instead of using only self-rating proficiency and fluency questionnaires I will be providing a concrete representation on participants' current foreign language proficiency level by adding both a standard language proficiency test in the FL and two verbal fluency standard tests. Even though FL proficiency was measured by Cavar & Tytus, (2018) using a standard proficiency test, that study focuses more on acculturation rather than purely the FLE phenomenon, by specifically testing a group of participants who learnt the L2 in the FL. However, in their study they did not directly compare the high and low acculturated participants. On the other hand, Miozzo et al. (2020), who also used an FL proficiency test only examined a highly proficient group of bilinguals. Their results challenge the effects of proficiency on the FLE (therefore the cognitive load hypothesis) however, the role of varying levels of language proficiency has yet to be determined with methodological rigor. Hence, this PhD aims to offer a significant and novel contribution to the existing literature on the FLE by systematically investigating the role of FL proficiency in decision-making by explicitly comparing low- and high-proficiency participants to determine whether proficiency moderates decision-making biases.

This study is methodologically innovative as no prior research has systematically compared proficiency levels within the FLE paradigm in a rigorous matter. Existing studies have typically examined self-rated proficiency for different proficiency levels or used rigorous proficiency measures only to ensure high proficient bilingual participants, without assessing how different degrees of proficiency might influence decision-making processes. By addressing this gap, this PhD research will provide a more nuanced understanding of the cognitive load mechanism underlying the FLE, contributing to broader theoretical debates in moral judgment research.

Hayakawa et al., (2017) emphasize however that the specific explanations behind why individuals opt for varied choices when using an L2 remain ambiguous or not fully understood and proposed more in-depth investigation between the two main factors: cognitive load that results from FL disfluency or reduced emotionality in FL processing due to the nature in which a FL was learned. Therefore, according to Hayakawa et al., (2016), to fully understand how the language we use affects the

choices we make, future research should be carried out on different linguistic backgrounds and skill levels, including factors like proficiency, learning methods, age at acquisition, and cultural connections. The first study of this PhD will attempt to shed light on the first of those variables, namely proficiency (leaving the door open to look at some of the other variables in an exploratory analysis of the same dataset on the FLE). Additionally, to explore the possible reduced emotional effect, participants' self-perceived emotionality on moral dilemmas will be measured (Wong & Ng., 2018) for experiment 1.

On a review by Białek & Fugelsang (2019) on Cavar & Tytus, (2018) study it is argued that the issue on the material of the study is that the authors failed to acknowledge that in four out of the six dilemmas used, the action of killing one to save various people also involved saving themselves; the person making the decision. Since these self-interest dilemmas require one to save themselves from death, the current study will refer to them as self-preservation which is defined by the Oxford Dictionary of English as the 'the protection of oneself from harm or death' and scenarios that do not require such action as non-self-preservation.

Białek & Fugelsang (2019) argue that using self-interest dilemmas cannot produce genuine utilitarian decisions and Kahane et al., (2018) add that they cannot be treated as valid examples of unbiased utilitarian decision making. Therefore, to investigate the factors above and the possible role of the two types on FLE the current study divided the moral dilemmas used into two categories: self-preservation (SP) and non-self-preservation (NSP) dilemmas and the statistical analysis was carried out treating such scenarios differently in order to compare the decisions on the two types.

Furthermore, another important aspect that will be addressed in this PhD is that previous research (see Circi et al., 2021 & Białek & Fugelsang, 2019 for reviews) used between-subjects designs where two groups of different participants were assigned to either the native or the FL condition. The problem with the between-subjects design is that it attributes possible differences between the participants tested (i.e "moral rules, thinking style and working memory" differ; p.684) whereas a within-subjects design will increase power as it mainly focuses explicitly on variances of the conditions and

the experimental control (Białek & Fugelsang, 2019). The current PhD will implement a within-subjects design, where the FLE will be tested on the same participants in an aim to provide a more accurate representation of the possible FLE by eliminating the uncertainty of results deriving by comparing data between people with different dispositional perspectives and qualities (Hu & Reiterer, 2009). In order to achieve this, a language mode (Grosjean, 1998) task is used to induce the target language, along with some comprehension questions to fully engage participants in the language mode task. Furthermore, in Experiment 1 of this PhD participants undertake the Oxford quick placement test (QPT) that will provide participants actual, current proficiency language level. The QPT is a standardized proficiency test assessing formal language skills, and that it corresponds to the Common European Framework of Reference (CEFR) for languages. Additionally, fluency is also measured in the first experiment using standard letter and category fluency tests. Last, participants' background information, such as foreign language acquisition and context use, will be examined once collected through the Language Experience and Proficiency Questionnaire (LEAP-Q) in order to investigate, in the exploratory analysis, what other background factors facilitate or moderate the FLE on moral judgement (Hayakawa et al., 2016).

Chapter 2 of this thesis delves into the theoretical underpinnings of the FLE , providing a comprehensive exploration of its conceptual framework. The rationale behind this investigation lies in the profound impact that language can have on cognition, particularly in decision-making processes. By examining existing research on the FLE, this chapter aims to elucidate the mechanisms through which processing information in an FL influences judgment and decision-making, emotional reduction in the FL and cognitive load. Furthermore, it seeks to establish the significance of this phenomenon within the context of moral judgement, setting the stage for the subsequent empirical investigation. Through a detailed analysis of relevant literature, Chapter 2 aims to lay a solid foundation for understanding the intricacies of the FLE and its implications for this thesis's research objectives.

Chapter 3 will outline the central methodology employed in experiments 1, and 2, offering a comprehensive overview of the common procedures utilized across these experiments. The rationale behind this approach is to ensure consistency and

comparability between the experimental procedures, facilitating a rigorous and systematic investigation of the research questions outlined in the thesis. By presenting the general methodology, this chapter aims to provide transparency and clarity regarding the empirical methods employed. Additionally, it will justify the chosen methodology by highlighting its suitability for addressing the specific hypotheses and research objectives outlined in the preceding chapters, based on the existing literature body establishing a robust foundation for the subsequent analysis and interpretation of the empirical findings.

Chapter 4 will present the first experiment, focusing on the impact of foreign language proficiency and foreign language fluency on the FLE. This chapter will systematically present the methods employed, including participant selection, experimental design, stimuli presentation, and data collection procedures. Following this, the chapter will provide a thorough statistical analysis of the results obtained from both experiments. Finally, a comprehensive discussion will be offered, which will delve into the implications of the findings, their alignment with existing literature. The two experiments and their respective analyses aimed to shed light on the first factor 'cognitive load' and its role on the FLE by employing robust measures.

Chapter 5 will introduce the second major experiment, which investigates the impact of language emotionality on the FLE. This experiment aims to elucidate the role of language emotionality by employing robust emotional measures, specifically the Positive and Negative Affect Schedule (PANAS-X). The chapter will provide a detailed account of the experimental design, including participant recruitment, stimuli selection, and procedures for eliciting and measuring emotional responses. Subsequently, it will present the statistical analyses conducted to evaluate the relationship between language emotionality and the FLE. The discussion section will critically examine the implications of the findings, considering how language emotionality influences decision-making in an FL context seeking to contribute valuable insights into the relationship between language emotionality and moral judgment.

Chapter 6 will offer a comprehensive general discussion of the findings in relation to the existing literature. This chapter will systematically review and synthesize the results

obtained from all experiments conducted in this thesis, considering their implications within the broader context of the FLE and what drives it. By comparing and contrasting the findings with previous studies, Chapter 6 aims to reveal the contributions and novel insights offered by the current PhD. Last, this chapter will look into the limitations encountered in this study while also provide potential avenues for future research and suggest refinement of methodology.

Chapter 7 will summarize the conclusions drawn from the examination of the FLE on moral judgment, with a particular focus on the influence of FL proficiency and emotionality. This chapter will obtain the main findings that derived from the two thorough empirical investigations conducted in this thesis (experiments 1 and 2), linking the findings back to the literature body by shedding light on the, previously explained, uncertainty on what drives the FLE on moral judgment proposed by Hayakawa et al., (2016).

Chapter 2. Theoretical Background

Millions of foreign language speakers are not aware that our choices depend on the language we use when making them. Moral judgment is what people perceive as ‘‘right’’ or ‘‘wrong’’ (Costa et.al. 2014) and the need for further research on how moral judgment and decisions are affected by foreign language (FL) use is dominant, as decision making in an FL can impact people all over the world. The reason is because communication and material processing in the International public policy involve foreign language use (Geipel et al., 2015). Moral judgment also refers to the process of evaluating actions or behaviours as right or wrong based on moral principles, values, or norms. It involves assessing the ethical implications of a situation and making a decision about the appropriate course of action (Haidt, 2001). This highlights the complex interaction between foreign language use, cognition and decision making, leading us to the intricate phenomenon of the FLE and the importance of examining the possible factors that mediate or moderate this in more depth (Hayakawa et al, 2016).

However, the mechanisms underlying the FLE are still unclear (Hayakawa et al, 2016; Circi et al., 2021). Does the FLE phenomenon occur due to the emotional reduction in the FL (because of the way it was acquired; non-emotional classroom context)? Or does the FLE occur due to the more deliberate processing created by lower L2 proficiency or disfluency? Hence, I emphasize the importance of reevaluating the perception of the FLE and its ramifications, particularly concerning moral decision-making by foreign language speakers (see also Wong & Ng, 2018). Hence, the current PhD will aim to focus on these two possible factors the ‘cognitive overload hypothesis’ and the ‘reduced emotionality hypothesis’ (Kirova & Camacho, 2021) by designing experiments that aim to weigh evidence towards one or the other, or both.

This theoretical Chapter delineates a theoretical account of the key terms and a rationale supporting the proposed research, which is to investigate the FLE on moral judgments and the role of foreign language proficiency and emotionality. It aims to analyse the theoretical underpinnings behind the FLE focusing on the role of language proficiency and emotional reduction in the L2 and the ways it effects moral decision

making based on previous studies on the field. The chapter unfolds across three distinct sections: Firstly, it elucidates key terms pertinent to the FLE. Secondly, it scrutinizes prior research, particularly focusing on the influence of foreign language proficiency and foreign language emotionality on the FLE. Lastly, it expounds upon the rationale underpinning the current investigation, weaving together insights gleaned from previous studies and theoretical frameworks. Additionally, the chapter extends to address the theoretical underpinnings that support the proposed research, thereby providing a comprehensive foundation for subsequent analysis and exploration.

2.1 The foreign language effect (FLE)

Decision-making refers to the cognitive process of selecting a course of action among several alternatives. It involves evaluating various factors, such as risks, benefits, consequences, and preferences, to arrive at a choice that aligns with one's goals or objectives (Hastie, 2010). The FLE is known as the systematic influence of an FL on decision making (Dylman, & Champoux-Larsson, 2020). For instance, the FLE refers to the consistent impact that using a language other than one's native tongue has on decision-making processes. When individuals operate in a non-native language, their choices and judgments can be notably different from those made in their native language (Hayakawa et al, 2016). Research (e.g. Costa et al., 2014) demonstrates that this effect can lead to changes in various aspects of decision making, including risk perception, emotional response, and moral reasoning.

The phenomenon is often attributed to the psychological distance created by using an FL, which can promote more rational and less emotionally biased decision making. Emotionally biased decision-making occurs when a person's emotions impact their choices, leading to a deviation from rational thinking. This influence often causes decisions to mirror the person's emotional condition at the time, rather than being based on a comprehensive and balanced evaluation of the facts (Lerner & Keltner, 2000).

In the framework of dual-process theory, cognitive operations are categorized into two distinct systems (Kahneman, 2011). The first, system 1, operates rapidly and automatically, requiring minimal mental effort. This system is primarily intuitive, drawing upon deep-seated experiences and emotions to guide decisions. It is

particularly effective in familiar contexts, such as when communicating in one's native language, where the processes are so well-rehearsed that they appear almost instantaneous and effortless (Kahneman, 2011). On the other hand, system 2 is characterized by its methodical and reasoned approach to problem-solving. This system is activated during tasks that demand attention and deliberate analytical thought, typically involving higher cognitive demands. Such conditions are often met in second language usage, where the lack of automaticity in language processing requires a more conscious and controlled cognitive engagement (Kahneman, 2011). This theoretical distinction underscores the variability in cognitive processing, depending on the complexity and familiarity of the task at hand. However, recent studies challenge the view that foreign language use consistently favours system 2 processes. For instance, Białek et al. (2020) found that high cognitive load in L2 contexts, can impair metacognitive monitoring (assessing and regulating one's own learning and understanding), which is essential for recognizing when analytical reasoning is necessary. This can result in an impulsive reliance on system 1, which is fast, intuitive, and more emotional, even in tasks where deeper reasoning is beneficial.

System 2 is also nuanced by findings from recent empirical research. For instance, Milczarski et al. (2024) thoroughly investigated the 'thinking more' hypothesis within a dual-process framework, examining whether using an FL leads to enhanced cognitive reflection and task engagement. Their extensive analysis, using data from over 1,700 participants across cognitive reflection and numeracy tasks, found no significant increase in system 2 engagement when tasks were performed in a FL. This suggests that the expected shift towards more rational or reflective thinking in L2 may not consistently occur as previously theorized, highlighting the complexity of cognitive processing across different linguistic contexts. Such evidence is crucial for refining our understanding of how language use influences cognitive systems, suggesting that the relationship between language and cognitive processing might be more variable than initially thought.

Additionally, L2 often reduces emotional salience, which can minimize biases and promote utilitarian decision-making (Hadjichristidis et al., 2015; Hayakawa et al., 2017). However, this emotional detachment can also suppress the emotional cues

necessary to trigger deeper reasoning processes (Białek et al., 2020). For example, when moral or ethical issues arise, the lack of emotional intensity in L2 may result in a failure to recognize the severity of certain moral judgements or the need for reflective thought, thus promoting utilitarian decision-making, particularly in moral dilemmas, where strong emotional responses often dominate L1 decisions.

Keysar, Hayakawa, and An (2012) suggest that the concept of psychological distance in the context of the FLE refers to the cognitive and emotional detachment experienced when individuals use a language that is not their native one. This detachment can create a sense of distance from the situation or decision at hand, leading to changes in cognitive processing and decision-making behaviour. The authors thus suggest that, using an FL can increase psychological distance from the specific decision at hand, making individuals less emotionally involved and more inclined toward rational deliberation. This occurrence is attributed to the reduced emotional resonance of words and phrases in a non-native language (Declerck & Reed, 2001; Harris et al., 2003; Jończyk et al., 2024), which diminishes the immediate emotional reactions when it comes to moral decision making. Therefore, psychological distance in an FL can promote a more analytical and objective approach to decision making by reducing the influence of cultural biases and ingrained emotional responses (see Costa et al., 2014). Individuals may rely more on the deliberate, systematic processing of information when operating in an FL, leading to decisions that are based on rational evaluation rather than intuitive reactions, e.g. make more utilitarian decisions; decisions that will benefit the greater good, regardless of potential harm (Costa et al., 2014).

In essence, while psychological distance refers to the overall detachment and reduced emotional impact experienced when using an FL, on the other hand, framing effects can make the same decision seem more or less attractive depending on whether it is framed positively or negatively (Tversky & Kahneman, 1981) based on the language used. Framing effects in foreign language use refer to how different presentations of information (positive or negative) impact decision-making in an L2 compared to a native one (Costa et al., 2014). As initially outlined by Keysar, et al., (2012), the FLE refers to a phenomenon where in bilingual individuals exhibit a diminished sensitivity to framing effects (Tversky & Kahneman, 1981); when making

decisions in a language other than their native one. As opposed to their proficiency in their native language the authors expand that the ‘‘automatic’’ (instinctive), processing is predominant in the native language. On the other hand, the possible emotional reduction that occurs (from lower proficiency and reduced fluency, when thinking in an FL), creates an emotional distance and causes the FLE when decisions taken in the FL vary from those taken in the native language. If foreign language use promotes reduced emotional mental patterns, it means that moral decisions taken in an FL are less impacted by emotional drive compared to the native language (Shiv & Fedorikhin, 1999).

Framing effects and psychological distance are interrelated in the context of the FLE, as framing effects pertain to how the presentation of information influences decisions, which can be altered by the psychological distance inherent in foreign language use (Keysar et al., 2012). This psychological distance creates a sense of detachment and reduces emotional involvement when using an L2, resulting in more analytical and less emotionally driven choices. As a result, this detachment can amplify the impact of framing effects when decisions are made in an FL (Keysar et al., 2012; Costa et al., 2014).

Psychological distance in foreign language use is also influenced by the cognitive load associated with using an FL (Hayakawa et al.,2017). For instance, using an FL is usually quite demanding because you have to process it more to comprehend the meaning in contrast to someone’s native language. This increases the mental effort; cognitive load, leading to a more detached, analytical approach, making people less emotionally biased (Hayakawa et al., 2016); meaning making decisions or judgments based on your emotions rather than on logical reasoning or facts (Smith, 2020). Whereas framing effects manipulate the context to elicit different emotional and cognitive responses without necessarily increasing cognitive load (Tversky & Kahneman, 1981). The relationship between psychological distance and the framing effect is significant, as both of these phenomena might actually be taking place at the same time as increased psychological distance encourages abstract thinking (Trope & Liberman, 2010) but the framing of information affects the perception and interpretation, often guiding people’s judgments and choices by emphasizing context

and presentation over the actual content (Tversky & Kahneman, 1981). Nevertheless, the connection between the two phenomena remains unclear (Trope & Liberman, 2010).

Psychological distance is linked to the cognitive load hypothesis as increased cognitive load can lead to more abstract thinking, which is a core aspect of psychological distance. When cognitive load is high, individuals are more likely to be more psychologically distant, influencing their perception and decision-making processes (Trope & Liberman, 2010). The cognitive load hypothesis posits that tasks requiring significant cognitive effort can exceed our limited cognitive capacity (Sweller, 1988). It is suggested that the cognitive resources used for foreign language processing leave fewer resources for emotional processing, leading to diminished emotional biases (Costa et al., 2014). This phenomenon has been particularly linked to decision-making scenarios, where individuals tend to make more utilitarian choices in an FL compared to their native language (Keysar et al., 2012). On the other side, the emotion- reducing hypothesis is linked to framing effects when using an FL because thinking in a non-native language can reduce emotional responses. This emotional distance can lead to more analytical and less biased decision-making, altering the way framing effects influence choices (Costa et al., 2014).

Thus, even though the two theories come across as interconnected by some means they are also distinct; cognitive load hypothesis focuses on how tasks can overload cognitive capacity (Sweller, 1988; Paas et al., 2003), whereas the emotion reduction hypothesis specifically examines the effect of FL use on emotional processing and responses (Hayakawa et al., 2016). In light of the fact that foreign language use typically dampens emotional intensity, regardless of FL proficiency, results in less emotionally driven decisions (Pavlenko, 2005). Hence, while it is argued that proficiency can moderate some detachment (Costa et al., 2014), yet emotional effects remain significant in influencing the FLE.

Beyond the fact that FL might require more cognitive effort, often foreign language speakers report different emotional responses in FL than the native language. An example for the above could be research that suggests that swearing in an FL may be easier due to reduced emotional impact (Dewaele,2010; Dewaele, 2004; Christofi, 2011), which challenges the cognitive load hypothesis. Typically, using an FL increases

cognitive load, making tasks more effortful and deliberative (Hayakawa et al., 2017). However, swearing might bypass this cognitive load due to emotional detachment (Dewaele, 2010). This discrepancy implies that certain automatic responses, like swearing, might not conform to the expected increase in cognitive load (Chen et al., 2022; Zhang, 2013). These findings imply that the cognitive load hypothesis may not fully account for all nuances of FL processing, particularly in emotionally charged or habitual expressions like swearing.

The above exhibits the notable distinction between the two main theories around the FLE which are thoroughly discussed in the next sections of this chapter: the emotion reduced hypothesis (Geipel et al.,2016; Keysar et al.,2012; Hadjichistidis et al.,2015) and the cognitive load hypothesis (Costa et al., 2014; Keysar et al.,2012; Cipoletti et al.,2016). The above leads to the conclusion that both cognitive load and the emotion-reducing hypothesis' role on the FLE on moral judgment need to be explicitly and separately explored in order to shed light to the aforementioned distinction and it creates an even more intriguing motive to thoroughly investigate the role of the cognitive load using explicit measures instead of self-rated proficiency. Nevertheless, Hayakawa et al., (2017) stress that apart from the emotional reduction in the FL there may be other factors that mediate or moderate the FLE on moral decision-making and stress that future research should thoroughly examine other influences that may influence the reduced emotional effect, such as how proficient a participant is (which experiment 1 will explore in the current thesis) and the method they acquired the FL.

The next section presents the theory behind the first possible parameter that could be driving the FLE. A theory review on the role of FL emotionality.

2.2 Emotion-reducing hypothesis in the FL

Emotions are complex psychological states that arise from a combination of physiological arousal, cognitive appraisal, subjective feelings, and behavioural responses. They can be characterized by a wide range of feelings such as joy, sadness, anger, fear, love, and surprise. Emotions play a crucial role in human experience, influencing our thoughts, decisions, actions, and interactions with others

(Lazarus,1991). According to Barrett (2017), the theory of constructed emotion (previously coined as conceptual act model of emotions, Barrett, 2006) defines emotions as dynamic, constructed experiences rather than fixed, biologically predetermined states. The theory challenges the traditional approach, which assumes emotions like anger or fear are universal and biological entrenched (e.g., Ekman, 1992), and instead proposes that emotions emerge through situated conceptualizations; a process in which the brain integrates sensory inputs, past experiences, and conceptual knowledge to create meaning. This approach highlights variability, emphasizing that emotions are not rigid categories, but flexible experiences influenced by cultural and linguistic contexts. Therefore, language plays a crucial role in shaping emotional experiences, as words guide attention to specific features of affective states, reinforcing socially learned patterns of emotion perception and expression.

This intricate link between emotions and language, influence both how we express ourselves and how we interpret the expressions of others. They shape the tone, nuance, and meaning conveyed through words, allowing for the communication of complex feelings and experiences. Moreover, language can also impact emotions by framing our thoughts and perceptions, shaping the way we experience and understand the world around us. According to Lindquist, Barrett, Bliss-Moreau, and Russell (2006), the accessibility of emotion words can significantly impact how quickly and accurately people identify emotional expressions. Their studies demonstrated that when individuals were temporarily unable to access an emotion-related word due to semantic satiation—a process that momentarily disrupts the meaning of a word through repetitive exposure—their ability to recognize facial expressions was impaired. This effect was observed even in perceptual tasks that did not explicitly require verbal labeling, suggesting that emotion words function as conceptual anchors that guide perception. These findings align with the linguistic relativity hypothesis, which posits that language influences cognition by structuring how individuals categorize and experience the world. By extending this hypothesis to the domain of emotions, Lindquist et al. argue that different languages may shape the way speakers from various cultures interpret emotional states. Instead, Lindquist et al., 2006 (see also Barrett, 2006; Gendron et al., 2014; Lindquist & Gendron, 2013) support a constructivist approach, in which emotions are shaped by

language and cultural context. As a result, individuals who speak languages with distinct emotional vocabularies may perceive and categorize emotions differently, highlighting the interdependent nature of language, cognition, and emotion perception. To sum up, according to constructionist perspectives (e.g., Barrett, et al., 2015; Lindquist, 2013; Lindquist et al., 2014, 2015; Wilson-Mendenhall, 2017), words serve as vital tools for organizing and unifying varied sensations into coherent concepts of distinct emotions. While constructionists acknowledge that sensations can be experienced independently, they emphasize that words function as a binding force, merging diverse experiences that might otherwise lack clear commonality into the structured understanding of discrete emotions (Lindquist et al., 2015). In essence, emotion words do far more than segment a spectrum of sensations—they integrate a wide array of experiences into unified emotional constructs (Lindquist, 2009).

Hence, constructionist models present a paradigm-shifting perspective on the nature of emotions, challenging traditional views that treat emotions as biologically hardwired entities with distinct neural signatures. Instead, the theory of constructed emotion (see Barrett, 2017), posits that emotions are not innate but rather arise from the brain's predictive processes that regulate the body in response to internal and external stimuli. This theory is rooted in neuroscientific evidence, rejecting classical models that assume emotions like fear, anger, and sadness have fixed neural circuits. Barrett continues that the brain does not react to stimuli in a stimulus-response manner but actively constructs emotions using past experiences, sensory input, and conceptual knowledge. The brain employs predictive coding—a process where it anticipates and interprets incoming sensory data rather than simply responding to it. This predictive nature of the brain is crucial for allostasis, the process of regulating bodily resources efficiently. Within this framework, interoception, or the brain's awareness of internal bodily states, plays a key role in shaping emotional experiences. Moreover, Barrett critiques traditional emotion theories that rely on the assumption that emotions have universal "fingerprints" in the brain, such as distinct facial expressions or physiological patterns. Instead, she presents empirical evidence demonstrating that emotions are highly variable and context-dependent, with no single neural signature consistently corresponding to any one emotion. The brain constructs emotional experiences

dynamically, based on cognitive concepts that help categorize sensations into meaningful experiences. The theory of constructed emotion offers an understanding of how emotions emerge—not as predefined biological responses but as contextually constructed experiences driven by the brain's predictive and regulatory mechanisms suggesting that emotions can be shaped and influenced by learning, culture, and context rather than being biologically predetermined.

For instance, the review paper by Dylman, Champoux-Larsson, and Zakrisson (2020) discusses how emotions are shaped by cultural contexts, with research showing that individuals are more adept at recognizing emotions within their own cultural group. Theories of emotion, such as universalist perspectives and dialect theories, seek to explain variations in emotional perception across cultures. The review paper also explores the relationship between language and emotion, emphasizing that bilinguals often experience greater emotional intensity in their native language. This phenomenon, known as the emotional distance hypothesis; in this PhD-emotion reduced hypothesis; suggests that second languages, typically learned in neutral contexts, evoke weaker emotional responses. Furthermore, decision-making processes would be influenced by language, with bilinguals demonstrating more rational, less emotionally driven decisions when using their second language (FLE).

Therefore, the language one grows up speaking significantly influences how emotions are perceived and expressed. This idea, often referred to as linguistic relativity, aforementioned, suggests that the structure and vocabulary of a language can shape the way individuals understand and articulate their emotional experiences. For instance, languages with rich emotional vocabularies may promote greater emotional awareness and precision in labelling feelings, while languages with more limited emotional lexicons may lead to broader or more ambiguous descriptions of emotions. Additionally, cultural norms regarding emotional expression and communication style, which are often embedded within language, further shape how emotions are conveyed and understood within a community (Boroditsky,2017). But what happens when we express our emotions in an FL?

When expressing emotions in an FL, individuals may encounter challenges related to linguistic and cultural differences. These challenges can affect the accuracy and intensity of emotional expression, as well as the ability to perceive and interpret emotions in others. Research suggests that emotional expression in a non-native language may be less spontaneous and authentic compared to expression in one's native language. Additionally, individuals may experience difficulties in finding appropriate words or cultural norms for conveying specific emotions, leading to potential misunderstandings or misinterpretations. Despite these challenges, expressing emotions in an FL can also offer opportunities for personal growth, cultural exchange, and the development of linguistic and emotional fluency (Pavlenko, 2012).

However, when experiencing emotions in an FL, individuals may perceive differences in the intensity or depth of those emotions compared to when experiencing them in their native language. The phenomenon of the FLE, extensively described in the previous section, is linked to the "emotional fluency effect" which suggests that emotional experiences can be influenced by the language in which they are processed (Pavlenko, 2006). People often report feeling less emotionally intense when expressing themselves in a non-native language. This effect may stem from the cognitive effort required to communicate in an FL, leading to a sense of emotional detachment or reduced emotional resonance (Greene et al., 2001; Harris et al., 2003; Keysar et al., 2012). Additionally, cultural and linguistic differences between the native and foreign languages can impact the way emotions are conceptualized and expressed, further contributing to variations in emotional experience across languages (Keysar et al., 2012).

In the past there has been considerable research demonstrating how our L1 is considered our most emotional language (Deweale & Pavlenko, 2001; Panayiotou 2006). It has been suggested that the FL has less emotional impact than the native language due to factors such as lack of emotional resonance or reduced emotional vocabulary in less proficient bilinguals (Pavlenko & Dewaele, 2002). Emotional words in the native language are stored at a deeper level of representation than words from an L2, for the reason that, words that express emotion in the native language are much more familiar to the individual since they have been used in numerous contexts and in

various ways (Altarriba, 2003). Whereas L2 exposure, in most bilinguals, is mostly limited to a (less emotional) classroom-based setting (Keysar et al., 2012). Dewaele (2008) adds that there are existing memory 'clues' for emotional words, generating a stronger semantic representation. However, emotion words in an L2 are not deeply encoded, as they have been practiced less and in fewer contexts. Hence, words in our L1 will activate more emotional associations, rather than the same words in an FL (Dewaele, 2008). This could be due to the psychological distance that might exist in the L2, as according to Pavlenko & Dewaele, (2002) less proficient bilinguals tend to avoid using emotional words in the L2, due to limited lexicon or lack of emotional resonance, in contrast to more proficient bilinguals.

This leads to the emotion-reducing hypothesis, as stated by which is based on research revealing that foreign language messages usually elicit less significant emotional responses in contrast to the L1 (Dewaele, 2004; Harris, 2004; Harris et al., 2003; Hsu et al., 2015; Iacozza et al., 2017). Studies revealed that expressing upsetting ideas in an L2 create a detachment on what bilinguals say (Geipel et. al., 2015). Therefore, people feel more comfortable to code-switch to their L2 in order to talk about awkward topics, as it is easier to discuss them for a longer time in their L2 in contrast to their L1 (Bond & Lai, 1986). This was also supported in the field of psychotherapy, where L2 use is preferred in order to maintain a feeling of emotional distance. (Altarriba & Santiago-Rivera, 1994; Schrauf, 2000).

Geipel et al. (2015) investigated how reasoning in a FL influences moral judgment, emphasizing the role of emotionality. Across four studies, they found that participants using a FL judged moral transgressions less harshly, especially in purity-related violations, such as consensual incest or eating a deceased pet. This effect was strongly tied to emotional attenuation—participants reasoning in a FL consistently reported weaker emotional reactions, which correlated with more lenient moral judgments. In Study 1a (38 German speakers; FL English, NL German), participants read and evaluated moral violation scenarios, such as incest between siblings, in either their NL or FL. Those using FL English rated these transgressions less severe than those using NL German. Study 1b replicated this finding with 60 Italian speakers (36 FL English, 28 NL Italian), confirming that this effect was not language-specific but rather a

generalizable FL phenomenon. Study 2 (78 Italian-English bilinguals; NL Italian) tested whether language proficiency influenced the FL effect. Participants evaluated a broader range of moral violations, including loyalty (e.g., betraying a friend), fairness (e.g., cheating in a game), and purity (e.g., using a national flag to clean a toilet). The findings confirmed that FL use consistently led to milder moral judgments, with the strongest effect for purity-based violations. Importantly, while higher proficiency weakened the effect, it did not eliminate it, supporting the idea that emotional detachment in a FL is not solely dependent on proficiency but also on the emotional resonance of moral norms. Study 3 (74 Italian speakers, 37 in FL German, 37 in NL Italian) directly tested the emotional attenuation hypothesis by measuring self-reported emotional distress alongside moral judgments. Participants who read and judged moral violations in a FL reported significantly weaker emotional responses, and this reduction in emotional intensity correlated with less severe moral condemnation. This provided direct empirical evidence that the FLE on moral judgment is driven by reduced emotional engagement, rather than cognitive complexity or linguistic factors. Their findings support the emotional reduced hypothesis suggesting that a FL reduces emotional salience, making individuals rely more on rational, deliberative reasoning rather than instinctive moral intuitions. The study highlights how language shapes moral cognition by altering the emotional weight of ethical decisions.

Moreover, a study by Vives, Costumero, Ávila, and Costa (2021) examined how bilinguals process and regulate emotions in their native (Spanish) and foreign (English) languages. It involved 26 participants, all unbalanced bilinguals (individuals who exhibit greater proficiency in their native language) with intermediate English proficiency and limited immersion in English-speaking environments. Participants engaged in an affect-labelling task while undergoing fMRI scanning, matching emotional stimuli across six experimental conditions involving emotional labelling, gender labelling, and shape matching as a control. Results revealed that emotional labelling in the native language reduced amygdala activation compared to affective matching, replicating prior findings on emotion regulation, whereas this reduction was absent in the foreign language. Interestingly, labelling emotions in a foreign language elicited greater amygdala activation than in the native language, suggesting that foreign language processing imposes additional cognitive demands that impair emotional

downregulation. The findings challenge theories such as distraction and symbolic conversion while supporting the reduction-of-uncertainty mechanism, which posits that weakened sensory and semantic processing in a foreign language heightens emotional uncertainty. The study underscores that language choice significantly impacts emotional experiences, with native language processing facilitating emotional regulation more effectively than a foreign language. This study, thus, provide additional neurophysiological evidence in favour of the emotional reduce hypothesis.

Additionally, using an FL has been demonstrated to decrease impulsive, heuristic decision-making processes, known as the affect heuristic, and encourage a more logical approach, thereby mitigating decision biases (Costa et al., 2014a; Keysar et al., 2012). The affect heuristic is described as an aspect of human thinking in which emotions act as signals to influence judgments and decisions, which entails swift and automatic emotional reactions (Slovic, 2007).

As stated in the beginning of this chapter, one of the mechanisms on reduced emotional hypothesis during moral judgments is the dual-process theory. According to Cipolletti et al., (2016) moral judgments can be driven by two different processes. The first type- system 1 is rapid, instant and non-conscious but the second type- system 2 takes longer and is systematic, more voluntary (Cipoletti et al 2016). The first type regards to emotional responses involved in decision making guided by emotions (where the L1 is usually involved) and the second type involves decision making in a more rational more neutral manner; less affected by emotions and thus more involved in FL decision making (Keysar et al., 2012; Cushman, 2013). The reason being this is that L1 is likely more emotional because it is learned in affect-rich contexts (parents, caregivers etc.) but the L2 is more neutral because it is learned in classrooms (likely to be affect-neutral). Therefore, the two types of processing are not mutually exclusive, it is rather a matter of degree. This statement suggests that the two types of cognitive processing described in dual process theory—system 1 and system 2—are not completely separate from each other; instead, they exist on a continuum. Therefore, the FLE points towards dissimilar forms of processes resulting in different moral judgments depending on the language used. (Cipoletti et al., 2016; Cushman, 2013). In other words, individuals can engage in both intuitive, automatic processing (system 1) and analytical, deliberate

processing (system 2) to varying degrees depending on the situation. Regarding the FLE , this implies that the FLE is indicative of different degrees of engagement of these cognitive processes depending on the language being used. Essentially, depending on whether one is using their native language or an FL, the balance between intuitive and analytical processing may shift, leading to variations in moral judgments and decision-making outcomes.

The role of an FL in the context of dual process theory can be understood through its influence on cognitive processes and decision-making. When individuals use an FL, they may experience cognitive shifts that affect both System 1 and System 2 processing. In terms of System 1, using an FL can disrupt automatic, intuitive thinking processes. This disruption occurs because individuals may need to exert more effort to comprehend and communicate in a language that is not their native tongue. As a result, the rapid, heuristic-based judgments typical of system 1 may be less pronounced in an FL context. On the other hand, foreign language use can also impact system 2 processing by necessitating more deliberate, analytical thought. Individuals may find themselves engaging in deeper cognitive processing to overcome language barriers and accurately convey their thoughts. This increased cognitive effort can lead to a more systematic and careful approach to decision-making, aligning more closely with the characteristics of system 2. Overall, the role of an FL within dual process theory involves both disrupting automatic cognitive processes associated with system 1 and promoting more deliberate, analytical thinking characteristic of system 2 (Costa et al., 2014).

Specifically, it is argued that instinctive emotional responses have an equally competitive role with more controlled mental responses (Greene, 2007). The reason behind this is that utilitarian moral judgments (judgments that favour the greater good despite an individual person's rights; (Costa et al, 2014) derive from a controlled mental procedure; known as the consequentialist response, whereas non-utilitarian moral judgements derive from instinctive emotional reactions; known as the deontological response (Greene, 2007). A utilitarian response refers to a decision-making approach that prioritizes the outcome or consequence of an action (Mill, 2012). In utilitarian ethics, the moral worth of an action is determined by its ability to produce the greatest

good for the greatest number of people (Mill, 2012). This perspective emphasizes the importance of maximizing overall happiness or utility, often through calculating the benefits and harms of different courses of action. On the other hand, a deontological response is rooted in deontological ethics, which emphasizes the inherent rightness or wrongness of actions themselves, rather than their consequences (Kant, 1785). Deontological ethics is based on moral rules, duties, or principles that guide behaviour, regardless of the outcomes (Kant, 1785). According to this perspective, certain actions are inherently right or wrong, irrespective of their consequences.

In terms of emotionally charged words, Zheng et.al., (2020) investigated the phenomenon of how bilingual speakers express taboo words (words or expressions that stimulate emotion) less easily in their L1 than in their L2. The aim of their study was to ascertain if electro dermal monitoring the physiological method used to measure changes in skin conductance, previously known as the galvanic skin response would indicate whether dissimilar degrees of autonomic reactivity will be generated when identical taboo expressions are presented in the participants' L1 or L2. To conduct the aforementioned comparison and explore the subsequent hypotheses, 32 bilingual speakers, whose L2 was English acquired after the age of 12, participated in the study. Their L1 was Turkish. The authors explored two potential explanations. The first hypothesis was that more intense skin conductance was expected when participants were exposed to taboo words in their L1 in contrast to analogous taboo words in their L2 (which participants learned subsequently in life). The second hypothesis predicted that L1 childhood emotional expressions, for example rebukes young children hear, would induce bigger physiological reactions, regardless if participants were aware of them or not. In order to investigate the hypotheses aforementioned the authors exposed the informants to a plethora of inducing words from five categories: "16 neutral (door), 16 positive (bride, joy), 16 aversive (disease, kill), 9 taboo (asshole, breast), and 7 reprimands of the type commonly spoken to children (Don't do that! and Go to your room!)" p. 567. Some of the inducing words were presented in the participants' L1 and some in their L2 either auditory (heard from a computer loudspeaker) or through visual presentation (read what was displayed on the computer screen). The participants' reactivity to each word was censored through fingertip electrodes while they also rated from a scale of 1 to 7 on how pleasant or unpleasant they considered each word/phrase.

When participants were exposed to words in their foreign language they also had to rate how familiar each word was to them using the same 1 to 7 scale. The results of the electrodermal monitoring revealed that taboo words in the L1 stimulate more responsiveness than in the L2, causing more anxiety when taboo words are expressed in the L1. It was also shown that auditory incentives provoked greater responsiveness than optical incentives. Furthermore, the skin conductance physiological responses demonstrated that emotion stimulating words can be expressed with greater ease in the L2 regardless if the taboo words were heard or seen during the experiment. What is more, it was revealed that emotional responsiveness to phrases from participants early years (such as "Shame on you!") mostly in their L1 than in their L2. The greatest scale difference between the L1 and the L2 occurred with taboo words or phrases from reprimands. Participants linked auditory reprimands in the L1 to childhood memories.

More recently, Kyriakou, Foucart, and Mavrou (2023) investigated the influence of language on moral decision-making, specifically focusing on how Spanish-English bilinguals responded to the emotionally charged footbridge dilemma. Participants were presented with the scenario in either their native language (L1, Spanish) or their second language (L2, English). Importantly, the study design included both a binary choice (yes/no) and a Likert-scale measure for moral permissibility as well as a more nuanced way to assess the emotional impact of the dilemmas. Participants were asked to decide whether to push a person onto the tracks to stop a runaway trolley and save five others and then to rate the moral permissibility of their choice. Open-ended questions followed, inviting participants to explain their reasoning and describe the emotions they experienced. The researchers applied inductive content analysis to categorize moral justifications into deontological (emotional) or utilitarian (rational) arguments. The results revealed that participants were more likely to make utilitarian decisions in L2 compared to L1, with L2 participants rating the decision to push the individual as more permissible on the Likert scale. Emotional vocabulary analysis showed that L1 responses contained significantly more high-arousal words, while L2 responses included more low-arousal words. Mediation analysis confirmed that the reduction in high-arousal emotional words in L2 mediated the effect of language on moral judgments, supporting the emotion reduced hypothesis. This finding suggests that the diminished emotional intensity in L2 facilitates more utilitarian reasoning. The

qualitative analysis of arguments further highlighted differences in reasoning. In the L2 condition, participants more frequently justified their decisions with utilitarian principles, such as prioritizing the greater good or minimizing harm, and reported a reduced emotional connection to their choices. Conversely, L1 responses included more deontological reasoning, with participants emphasizing the intrinsic value of life, feelings of guilt, and concern for moral and legal consequences. Some L1 participants expressed deep emotional distress, citing their inability to carry the burden of causing harm. By comparing high- and low-arousal words and conducting mediation analysis, the authors provided evidence that support the emotional reduced hypothesis. The authors suggest future studies should explore additional moral dilemmas with varying levels of emotionality, incorporate physiological measures of emotional arousal, and investigate the role of individual differences in bilingualism, such as proficiency and linguistic context, to gain further insights into this phenomenon. Nonetheless, these more nuanced methods of assessing participants' experiences provide evidence that individuals exhibit reduced emotional engagement during moral decision-making, thereby lending support to the emotion-reduction hypothesis.

A study by Hadjichristidis et al., (2015), examined the extent to which foreign language use affects judgments of risk and benefit. What is more, the authors sought to test the 'foreign language hypothesis' supposition that when using an FL, a more positive effect will be activated on stimulus in contrast to native language use. Based on the pre-existing theoretical underpinnings the authors hypothesized that the 'affect heuristic' will affect risk and benefit judgments. Positive affect is linked to safety (high-benefit and low-risk), and negative affect is linked to fear (low-benefit and high-risk) (p.1). Italian-English participants had to evaluate certain tasks as a risk or as a benefit to the Italian community in either their native or foreign language. The results revealed that the authors correctly anticipated that foreign language use will produce more positive responses on benefits and less risk judgment in accordance with the affect heuristic. Informants were asked to provide the degree of how they felt (positive or negative), on a 5-choice scale, for each of the questionnaire items in either their native (Italian) or foreign language (English). The findings from the statistical analysis on feelings revealed that foreign language use reduced negative emotions on judgment of risk and intensified positive ones on benefit, supporting the underpinnings of the FLE.

The authors clarify that foreign language use can have an impact on very important decisions due to reduced negative feelings, giving a feeling of a more positive impression on a thread influencing the actions taken. Last, they state that in contrast to similar studies they also reveal that foreign language use can also affect judgments through feelings. This could be connected to the reduced emotionality hypothesis as reduced negative emotions in the L2 could lead to more utilitarian moral decisions towards harmful situations.

It is believed that judgement on moral dilemmas derives from our deep values, therefore as long as there is a comprehension of a moral dilemma, the language that is being presented in shouldn't be an issue (Costa et al., 2014). However, thinking in a FL helps reduce biases in decision-making (Keysar et al.,2012). Less emotional scenarios promote utilitarian responses in both native and foreign languages, however the more emotional a dilemma is, the more utilitarian (a choice that will benefit the greater good) the responses are in a foreign language compared to the native language (Costa A, et al.2014).

To gain further understanding on the role of emotions in moral dilemmas, Shin and Kim (2017) investigated how the use of a foreign language (L2) compared to a native language (L1) affects moral reasoning and emotional engagement based on psychological distance. In their first experiment, 161 Korean participants evaluated four moral dilemmas, presented in either Korean (L1) or English (L2), to assess the impact of language on moral decision-making. The dilemmas included two personal scenarios, which involved direct emotional engagement (e.g., physically harming one person to save many), and two impersonal scenarios, which required more abstract and deliberative reasoning (e.g., flipping a switch). Participants responded to each dilemma by choosing either a utilitarian action (focused on maximizing outcomes) or a deontological action (focused on moral principles). The findings revealed that in personal dilemmas, participants made significantly more utilitarian decisions when using L2 than L1, suggesting that L2 use reduces emotional involvement. Conversely, no language-related differences were observed in impersonal dilemmas, where emotional engagement was less critical. This suggests that L2 facilitates a cognitive, rather than emotional, approach to morally complex decisions. The second experiment

aimed to explore whether L2 affects self-bias, a phenomenon where individuals prioritize self-related over other-related information. A smaller group of 26 participants completed a computerized shape-label matching task, where geometric shapes were paired with labels such as "me," "friends," and "others." Reaction times (RTs) were measured as participants identified whether the pairings were congruent (e.g., "me" matched with a specific shape) or incongruent. The results showed a robust self-bias effect in L1, with faster RTs for self-related stimuli compared to others-related stimuli. However, this self-bias effect was significantly weaker in L2, indicating reduced emotional salience when processing self-related information in a foreign language. Interestingly, while RTs highlighted the diminished self-bias effect in L2, error rates did not differ significantly between L1 and L2, suggesting that reaction times are a more sensitive measure of emotional processing in such tasks. These findings collectively suggest that using a foreign language increases psychological distance by reducing emotional engagement. The study emphasizes how L2 use impacts personal moral dilemmas, leading to utilitarian decision-making, and diminishes emotional attachment to self-related stimuli, as seen in the weaker self-bias effect. These results highlight the distancing mechanism of L2 in morally and emotionally charged contexts.

In addition, Brouwer's (2021) study explored the FLE by examining bilinguals' moral decision-making in their native (Dutch) and foreign (English) languages on personal and impersonal dilemmas) but focused on investigating whether the FLE is influenced by emotional intensity (personal vs. impersonal dilemmas) and the mode of presentation (listening vs. reading). A total of 154 highly proficient Dutch-English bilinguals participated in the experiment. Participants were exposed to moral dilemmas categorized as either personal (emotionally intense—e.g., the *footbridge dilemma*—or impersonal and less emotional—e.g., the *switch dilemma*) in either the native or foreign condition. These dilemmas were presented in two modalities: a listening task, where participants heard the scenarios read aloud, and a self-paced reading task, where participants progressed through the text on a screen at their own pace. The reading and listening tasks were designed to closely mirror one another. In the listening condition, participants wore headphones and heard the dilemmas read by a proficient bilingual speaker to maintain consistency across languages. In the reading condition, the task mimicked the linear structure of auditory information by presenting sentences

incrementally on separate screens. After processing each dilemma, participants judged the appropriateness of the proposed action by selecting either “yes” (a utilitarian decision) or “no” (a deontological decision) using designated buttons. Responses were collected under time constraints to standardize decision-making processes across participants. Language emotionality was inferred from established ratings in prior research, where moral dilemmas were categorized as personal (highly emotional, averaging 6.5/7) or impersonal (less emotional, averaging 3.6/7), based on their ability to evoke emotional responses; grounded in validated empirical evidence (Koenigs et al.,2007). The results of Brouwer’s (2021) study revealed a clear FLE for personal dilemmas, with bilinguals making significantly more utilitarian decisions in English than in Dutch. No such effect was found for impersonal dilemmas, suggesting that the FLE is driven by the emotional intensity of the scenario. Furthermore, participants exhibited a stronger tendency for utilitarian decisions during listening tasks compared to reading tasks, regardless of language. This difference in modality likely reflects the reduced cognitive load associated with auditory processing, potentially aided by prosodic cues in spoken language. The findings underscore the context-dependent nature of moral decision-making, influenced by both language-induced emotional detachment and task-specific cognitive demands. These results contribute to the growing body of evidence that foreign language processing fosters controlled, rational decision-making, particularly in high-stakes, emotionally charged scenarios.

Based on the above studies experimental evidence demonstrates that FL decreases emotional reactivity in contexts such as moral judgments, risk assessments, and self-related stimuli processing. Electrodermal studies (e.g., Zheng et al., 2020) reveal that taboo words elicit stronger emotional responses in the L1 compared to FL. Similarly, language emotionality impacts moral dilemmas differently depending on emotional intensity. For instance, Kyriakou et al. (2023) found that participants using FL were more likely to make utilitarian decisions, mediated by lower use of emotionally charged words. Also, the studies above emphasize that FL use tends to reduce emotional engagement, fostering more utilitarian decisions, especially in emotionally intense scenarios (e.g., personal moral dilemmas like the *footbridge dilemma*). In contrast, impersonal dilemmas, which are less emotionally charged, do not induce the FLE (Shin & Kim,2017; Brouwer, 2021). This was also observed in studies by Wong, & Ng,

(2018) and Čavar & Tytus (2018) where no significant utilitarian responses on impersonal (less emotional) moral dilemmas were recorded. Hence, the evidence presented above supports the emotion-reduced hypothesis (which suggests that reduced emotional intensity in L2 facilitates rational and utilitarian reasoning). Furthermore, since the FLE is usually only observed in high-stakes emotional moral scenarios, or emotionally charge words/events, in this PhD I will specifically focus on personal moral dilemmas.

However, language emotionality on the FLE has been mostly elicited using a single question, e.g., rate how upset you feel after reading the dilemma (Geipel et al.,2015; Wong & Ng., 2018). Yet, specific emotions such as anger and empathy (e.g. Cordellieri et al., 2020) are highly probable to be induced when it regards to moral dilemmas, thus the role of different types of emotions should be thoroughly explored on the FLE (experiment 2 of this PhD).For example, an important aspect that needs to be considered is that moral dilemmas often evoke negative emotions such as guilt, anxiety, and sadness. Individuals facing ethical decisions may experience emotional distress due to conflicting moral principles or potential consequences of their choices (Tangney et al., 2007). Guilt arises from perceived violations of personal or societal moral standards, leading to self-reproach. Anxiety accompanies uncertainty about outcomes or pressure to adhere to ethical principles amidst conflicting values. Sadness may stem from the inevitability of harm or loss associated with difficult decisions, especially when no clear solutions exist (Haidt, 2003). According to Koenigs & Tranel (2007) emotional engagement (in the L1) plays a crucial role in moral cognition, as evidenced by individuals with ventromedial prefrontal cortex (vmPFC) damage, who exhibit a heightened tendency toward utilitarian moral judgments. This shift occurs because emotional deficits diminish the natural aversion to causing harm, leading to a more outcome-driven, rational approach to ethical decision-making. Hence, if a FL has a similar effect of minimizing the intensity of emotions, investigating the specific emotions induced by moral dilemmas is vital for addressing the emotional aspects of ethical decision-making across various domains.

For example, an empirical study by García-Palacios et al., (2018) examined the impact of the FLE on a strong emotion which is fear. Fear is a strong negative feeling

that we communicate through language. Particularly, using a fear conditioning experiment in either participants' native or foreign language, the authors investigated the possibility that foreign language can affect the way fear is acquired. The authors conducted an instructed fear experimental protocol wherein participants received verbal instructions indicating that two distinct stimuli could either signal the potential presence of a threat stimulus or its absence. To gauge emotional reactivity, changes in participants' pupil size and electrodermal activity were recorded. The results of the study provide evidence that language context could potentially influence specific verbal conditioning processes, precisely that a foreign language may reduce fear conditioning because of the psychological distance that occurs with foreign language use in foreign language environments.

Another important factor that could be related to reduced emotionality in the FL proposed by Hayakawa and Keysar (2018) could be that FL reduces mental native-ness of the language that may contribute to the vividness of mental imagery. Mental imagery is reduced with the use of a foreign language. The role of mental imagery is considerable in shaping our emotions, thoughts, and even actions (Hayakawa & Keysar, 2018), the native language activates more lively mental imagery which is surpassingly decreased when using a foreign language, because of less access to sensory memories, which are very important for cognitive imagery of the past (Schacter, Addis, & Buckner, 2007). The results from Hayakawa and Keysar's (2018) study revealed that visual imagery is reduced in a foreign language in contrast to the first language. Additionally, when it comes to moral decisions on dilemmas (for example to sacrifice one life to save five), it has been revealed that more utilitarian responses are produced both in the FL (Costa et al., 2014) and when mental imagery is reduced (Amit & Greene, 2012). The above reveals that they may result from a shared mechanism.
It appears that individuals are less inclined to support the utilitarian act of pushing the man off the bridge if they can vividly visualize the scene (Amit and Greene, 2012). Conversely, when visual interference disrupts the ability to vividly picture the scenario, individuals become more willing to opt for the utilitarian, but emotionally difficult, choice of sacrificing one life to save many. In moral dilemmas, individuals using a foreign language tend to be more willing to sacrifice one person to save five, sometimes even twice as likely compared to those using their native language (Costa, et al., 2014).

In Hayakawa and Keysar's (2018) last experiment participants are presented with the *footbridge dilemma*, where they are asked to state whether they would push a man from a bridge on a train in order to prevent the train from killing 5 other people, and to also rate the vividness of the scenario (rate the clarity of their mental images related to the potential sacrificial victim and the five individuals who could potentially be saved, and the overall scenario by choosing options such as "no image" to "absolutely clear image. The results revealed reduced mental imagery on moral choice in their FL, explaining to some extent the role on the FLE. What is more, the authors also argue that foreign language use alters our mental images affecting decision making. Furthermore, the reduction of mental imagery partly justifies the foreign language effect when making moral choices because more lively imagined outcomes seem more probable to happen (e.g. Sherman, Cialdini, Schwartzman, & Reynolds, 1985). Weaker mental imagery concerns possible undesirable consequences affects risk perceptions and the willingness to take risks (Traczyk et al., 2015). Therefore, if foreign language reduces mental imagery, according to Hayakawa and Keysar (2018), it could explain formerly detected effects on how foreign language reduces risk perceptions (Hadjichristidis et al., 2015) and increases willingness to take risks. (Costa et al., 2014; Keysar et al., 2012). However, Montero-Melis and Ostarek (2020) challenge the findings of Hayakawa and Keysar (2018), arguing that the key results of their study are better explained by reduced language comprehension in a foreign language rather than by diminished mental imagery. The role of foreign language proficiency and the FLE will be explicitly discussed in the next section of this chapter, stating the importance of investigating the role of foreign language proficiency in more depth.

To investigate the reduced emotionality hypothesis at the level of event perception rather than the more complex process of decision-making, Woumans and Duyck (2020) examined how using a foreign language influences perceptions of crime severity, specifically assessing whether emotional engagement is reduced during foreign language processing. The study included 558 Dutch-speaking first-year psychology students at a university, all late learners of English, who acquired the language primarily through formal education. Participants were randomly assigned to one of four conditions based on language (native Dutch or foreign English). They assessed the severity of four distinct murder scenarios, ranging from crimes of passion to

premeditated greed, by rating them on a scale from 0 (not severe) to 9 (most severe). The scenarios were translated and back-translated to preserve meaning and emotional nuance. The results showed that participants rated crimes described in FL as consistently less severe than those presented in NL. Emotionality was inferred through these ratings, with lower severity in the foreign language condition suggesting emotional attenuation. The authors hypothesized that this effect could stem from reduced vividness of mental imagery and decreased emotional intensity during foreign language processing. These findings align with broader research on the foreign language effect, where analytical reasoning is prioritized over emotional responses in non-native languages. Interestingly, the study underscores significant implications for multilingual judicial contexts, where the use of a foreign language may inadvertently lead to more lenient judgments, potentially influencing sentencing and legal outcomes. In the context of this PhD, this study further supports the idea that when bilinguals process information in their FL they tend to judge events as less emotionally charged than in their first language, providing support for the emotion-reduced hypothesis.

In addition, Dylman & Bjärtå, (2019) aimed to investigate whether processing text in an L2 could reduce the experience of negative emotions. Participants read text excerpts in both their native language (Swedish) and their L2 (English), drawn from fiction books available in both languages. They rated their distress levels and answered questions about the text in either language, with the order of tasks counterbalanced. The design allowed for exploration on how processing negative content in the L1 versus L2 affected distress levels. The key question was whether reading negative content in the native language and then processing it in a foreign language could lessen the experience of distress. The results revealed that despite being highly proficient in English as an L2, participants' dominant language remained Swedish (as they demonstrated better accuracy in answering questions about the text when it was presented in Swedish (L1) compared to English (L2), and when responding in Swedish compared to English). However, the most significant discovery was that participants reported lower distress levels after answering questions about negative texts when the responses were in English (their L2), in contrast to when responses were in Swedish, which resulted in higher distress ratings. Specifically, when processing the text originally read in the L1 in the L2 that this effect occurred. The results support the hypothesis that L2 use might

diminish distress levels following a distressing event, in this case, reading text excerpts with distressing negative content. However, the results do not definitively differentiate between two potential explanations: whether higher cognitive load during L2 processing or weaker emotional connections in L2 underlie this effect. In connection to this, the authors recognize that relying on self-reported assessments of L2 proficiency rather than standardized language proficiency tests could be a limitation in their study. Subsequent research should consider comparing outcomes between these two types of L2 proficiency measures or even utilizing both methods simultaneously.

To sum up, the studies above collectively demonstrate how the FLE influences emotional processing and moral decision-making. They highlight that emotions in moral dilemmas have mostly been assessed using a single question, despite the significant role of specific emotions like anger and empathy. Various empirical findings suggest that FL use reduces emotional intensity, which in turn affects moral reasoning, fear conditioning, and risk perception. Hayakawa & Keysar (2018) revealled that the diminished emotional response in a FL may stem from reduced mental imagery, leading to a more utilitarian decision-making style. This aligns with Costa et al. (2014) and Amit & Greene (2012), who found that individuals using a foreign language are more likely to make utilitarian moral choices due to weaker mental visualization of the consequences. Similarly, Woumans & Duyck (2020) demonstrated that crimes described in a foreign language are perceived as less severe, suggesting a broader emotional attenuation effect. Moreover, the reduction in mental imagery also explains why individuals take greater risks when using a foreign language, as observed in Hadjichristidis et al. (2015). However, Montero-Melis & Ostarek (2020) challenge these interpretations, arguing that reduced comprehension rather than weaker imagery could be responsible for the observed effects. Collectively, these studies highlight the critical role of emotional attenuation in FL processing, while also raising important questions. Although some research suggests that reduced mental imagery may contribute to the FLE, it is also possible that this reduction stems from lower language proficiency—for example, due to conceptual representations not being fully consolidated. As noted by Montero-Melis and Ostarek (2020), as well as Dylman and Bjärtå (2019), further investigation into factors such as language proficiency is essential for deepening our understanding of the mechanisms and theoretical foundations underlying the FLE.

In the current PhD the role of the emotion-reducing hypothesis was examined by measuring participants' self-perceived emotionality on each of the moral dilemmas (Wong & Ng.,2018) with an emphasis on the two groups of dilemmas aforementioned, the SP and NSP dilemmas-in the first experiment and the 2nd experiment explicitly examined the possible emotional reduction by rigorously measuring the emotional state of participants before and after they have been exposed to moral dilemmas, in both languages. Therefore, in order to investigate the role of emotionality on the FLE in the second experiment rigorous measures, PANAS-X, were employed. The PANAS-X assesses these emotions through self-report, asking participants to rate the extent to which they have experienced various emotional states within a specified time frame. This instrument is valuable in research and clinical contexts, providing insights into emotional well-being, mood states, and psychological functioning (Watson et al., 1988).

This section, of the theoretical background chapter, delved into the concept of emotionality on the FLE and on moral judgment, thoroughly examining how the use of a foreign language can influence decision-making processes, especially within moral contexts. Research indicates that individuals are more likely to make utilitarian decisions; those based on outcomes rather than deontological principles; when using a foreign language compared to their native tongue (Costa et al., 2014). This phenomenon is attributed to the diminished emotional resonance of a foreign language, which reduces emotional responses and promotes more detached and analytical reasoning (Keysar et al., 2012). By analysing various studies, the chapter demonstrates how the FLE can significantly alter moral judgments by decreasing emotional interference (Hayakawa et al., 2017) and how further exploration on the exact role of language emotionality on the FLE could shed light on to the current uncertainty on what drives the FLE (Hayakawa et al., 2016).

In a study on how regularly bilinguals use emotion words in their L2, one of the fundamental reasons that affect the use of emotion words in the L2 is language proficiency (Dewaele & Pavlenko, 2002). Less proficient learners have a tendency to intentionally avoid the use of emotion words, due to limited vocabulary (disfluency) or the absence of emotional significance, whereas highly proficient speakers use emotional words in their L2 more often. Additionally, bilinguals that have a very similar level of

proficiency in their L2 as in their L1 demonstrate complete emotional word processing in emotional Stroop tasks (i.e. participants identify the colour of words, with negative words like "fear" (Eilola, Havelka, & Sharma, 2007).

Following this, the next section will provide an in-depth exploration of the background theory of the cognitive load hypothesis, the second potential factor that could possibly drive the FLE (Hayakawa et al., 2016), elucidating its interaction with and influence on cognitive processes across the FLE on moral judgment and decision-making.

2.3 The cognitive load hypothesis

The idea that in a foreign language you have to put more mental effort is known as the cognitive load hypothesis (e.g. Sweller,1994). The cognitive load hypothesis offers a theoretical framework that connects human cognition with instructional design (Sweller,1988). This theory delves into the allocation and utilization of cognitive resources during the processing and acquisition of information, drawing insights from an evolutionary perspective on the formation of human cognition (Paas & Sweller, 2012). Knowledge can be classified into two categories: primary and secondary knowledge (Geary, 2008). Primary knowledge, like native language acquisition, is effortlessly acquired by humans without explicit instruction (Tricot & Sweller, 2014). In contrast, secondary knowledge, for example foreign language learning requires conscious effort (Sweller et al., 2007).

According to Anton et al., (2020) an L2 is more difficult to use, therefore, more cognitive resources are required, increasing the mental effort needed to process information in contrast to the native language. As most of the cognitive resources would be engaged in understanding the FL then less resources will be available for the processes of decision-making. Research indicates that understanding a foreign language (FL) places considerable cognitive demands on individuals, significantly more so than processing one's native language. This is because FL comprehension requires managing unfamiliar linguistic structures and vocabulary, which intensifies cognitive load (Majerus, 2013; Majerus et al., 2019). Neuroimaging studies reveal that different neural pathways are activated during FL processing, further complicating the task and

increasing mental effort (Guidali et al., 2019). Additionally, cognitive control mechanisms are frequently engaged to handle less familiar linguistic inputs, further increasing the cognitive load (Chang, 2012; Clahsen & Felser, 2006). Further, studies on text-level comprehension reveal that FL learners face challenges in making sense of implicit information and maintaining coherence, adding to their cognitive burden (Foucart et al., 2016).

Costa et al., (2014), state that when cognitive load is high, individuals are more likely to rely on heuristic biases in their decision-making processes as well (Benjamin et al., 2006; Whitney et al., 2008; Forgas et al., 2009). Heuristic biases are cognitive shortcuts people use to make swift decisions, often causing deviations from logical or optimal outcomes. In the context of the FLE, the use of an L2 can alter these biases, typically reducing emotional impact and changing cognitive processing relative to one's native language (Costa et al., 2014). When cognitive load takes place in the FL, that requires analysis and effort, intuitive answers that would have taken place in the L1 processing (automatic route) cannot be checked or controlled in the FL Kirova & Camacho (2021). Therefore, similarly to Hayakawa et al., (2017), the authors continue that the increased cognitive load when reading in a FL leads to responses being affected by cognitive 'shortcuts'. A cognitive shortcut, also known as a heuristic, is a mental shortcut that allows people to make fast decisions by simplifying complex problems (Kahneman, 2011). Humans often rely on these mental shortcuts, known as heuristics, to make quick judgments and decisions, especially when faced with uncertainty or limited cognitive resources. These heuristics, while efficient, can lead to systematic biases and errors in judgment. Kahneman (2011) distinguishes between two thinking systems: system 1, which is fast and intuitive, and system 2, which is slower and analytical. While heuristics can be helpful in conserving cognitive effort, they can also result in cognitive biases when applied without careful consideration of the context.

But how is the cognitive load linked to foreign language proficiency? The less proficient someone is in a foreign language, the greater the cognitive load they experience when using it (Sweller, 1988). Foreign language proficiency refers to an individual's competency and mastery in a language other than their native tongue,

encompassing skills in listening, speaking, reading, and writing (Johnson, 2019). The FL proficiency scale spans from basic understanding to advanced knowledge and is commonly assessed using standardized tests that correspond to the CEFR scale for languages (Johnson, 2019). According to this, supposedly the more proficient an FL user is the faster it is to process and understand the FL and subsequently the less the cognitive load (Schweizer, Craik, & Bialystok, 2013; Kroll & Bialystok, 2013; Costa et al., 2014).

2.3.1 The role of second language proficiency on the FLE

One of the major studies that supports the assumption that the FLE occurs due to low foreign language proficiency is the one by Costa et al. (2014). This study revealed that lower L2 proficiency leads to more utilitarian choices (e.g., willing to sacrifice one life to save 5) in contrast to choices made by more proficient L2 foreign language speakers (Costa et al., 2014; Keysar et al.,2012; Cipoletti et al.,2016; Degner et al., 2012, Dewaele, 2010).

Specifically, Costa et al. (2014) conducted two experiments to examine the effect of FLE on moral decision-making, hypothesizing that processing moral dilemmas in an FL would increase utilitarian choices due to reduced emotional engagement. In experiment 1, participants from multiple linguistic backgrounds, including English, Spanish, Korean, Hebrew, and French speakers, were presented with the *footbridge dilemma*, a classic moral scenario in which one must decide whether to push a man off a bridge to stop a train from killing five others. Participants were randomly assigned to complete the task in either their native or an FL. The findings showed that those using an FL were significantly more likely to choose the utilitarian option, sacrificing one person to save five, compared to those responding in their native language. To ensure that these results were not driven by random responding or cognitive difficulty rather than emotional detachment, experiment 2 introduced the *switch dilemma*, a variation in which participants could pull a lever to divert a train onto another track, killing one person instead of five. Unlike the *footbridge dilemma*, this version is considered less emotionally aversive, as it does not require direct physical harm. The results revealed that the FLE was present only in the emotionally intense *footbridge dilemma* but not in

the less emotional *switch dilemma*, supporting the hypothesis that emotional reduction, rather than cognitive strain, drives the effect. Additionally, the authors examined the role of language proficiency, showing that the FLE was weaker among more proficient speakers, suggesting that greater fluency increases emotional grounding in the FL. The authors argue that higher proficiency allows individuals to develop stronger emotional associations with the FL, making their moral judgments more similar to those made in their native language. The findings indicate that while late learners of an FL typically exhibit the FLE, those with higher proficiency experience a diminished emotional detachment, thereby reducing the shift toward utilitarian reasoning. Costa et al. (2014) suggest that as proficiency improves, individuals may internalize moral and emotional constructs more deeply in their second language, gradually closing the gap between native and non-native decision-making patterns. However, the conclusions of this study are based on self-rating scale proficiency which according to Tomoschuk et al. (2019), even though extensively used in research on bilingualism (Li, Sepanski & Zhao, 2006), because of data collection ease (participants rating how proficient they are in a FL from a scale of i.e 1-7, 1-10) this method has some major drawbacks. Self-ratings could be subjective and inconsistent with the actual participants' level of proficiency. Therefore, more objective proficiency measures maximize proficiency scores accuracy and will be more reliable across studies (Tomoschuk et al., 2019).

Hence the current study aims to shed light on the above by requiring participants to take a standard placement test (Oxford QPT) that will provide participants' actual, current proficiency language level in order to provide more solid results to examine and contrast with self-rated proficiency. Montero-Melis & Ostarek, (2020) stress out the importance of language proficiency and reduced foreign language comprehension on FLE. Language proficiency is a very important aspect that should be very carefully taken into consideration because when there is a lack of understanding/uncertainty in some words in the FL and participants are unsure of something, they will rate towards the middle of the Likert scale (Douven, 2018). Likert scales are frequently used in research to measure the degree in which a respondent agrees with a given statement and a mid-point is normally included (Likert, 1932).

Wong & Ng, (2018) supported that the FLE is reduced, or there is no FLE, when there is high proficiency, and the FL becomes an L2. The authors investigated the FLE

on morality from the perspective of early bilinguals where the 'nativeness' of the L1 is not as dominant. They sought to validate the FLE of early bilinguals' languages on moral decision making and reveal how dominance in the L2 can affect moral decisions by leading to more utilitarian decisions in their L2 .86 early English-Chinese bilinguals took part in the experiment. Participants randomly performed a moral dilemmas task consisting of personal (they had to hypothetically commit an action themselves e.g., push a stranger on train tracks to save 5 lives; *footbridge dilemma*) and impersonal dilemmas (flip a control switch to kill someone and save 5 lives; *trolley dilemma*) in either English or Mandarin Chinese followed by a language background questionnaire presented in English. In order to test participants' bilingual language background, the authors used the Bilingual Language Profile BLP (Birdsong et al., 2012) an instrument valuating language dominance and experience (in the 2 languages) through self-reports (testing language history, use, proficiency, and attitudes). Five sets of scenarios (*Burning building, crying baby, organ transplant, shark attack, and trolley/footbridge* dilemmas) requiring either a personal or an impersonal action were adapted from a moral dilemmas list from the study of Christensen et al. (2014) and were translated to Mandarin Chinese in order to be used for the moral dilemmas task. In all scenarios the participants had to choose between sacrificing a single life to save 5-11 lives. They read the scenarios and then the probable choice and impact before choosing from a scale of 1 to 7 (utilitarian scale: a choice that will benefit the greater good) how likely it would be that they commit the action in the given scenario,. In addition, participants had to rate how upsetting each scenario felt. The results of the findings revealed that the FLE has an impact on early bilinguals' moral decisions but depends on language dominance. The authors explain that more dominant in the tested language reveals more difference between choices in personal and impersonal dilemma responses. More difficult personal choices are easier to make in the non-dominant language. They argue that the more confident participants felt toward the language, the less likely it was to choose a utilitarian response for a personal action in contrast to an impersonal action. The authors also revealed that stronger emotional stimulation in the personal choice moral dilemma corresponded with more utilitarian results which was not related to the participants' language dominance (See previous section on the emotion reduced hypothesis). However, this study examined language proficiency based on participants' self-

perceived foreign language proficiency which could differ from their current/actual language level (Tomoschuk et al., 2019).

Nevertheless, even though proficiency was previously measured by Cavar & Tytus (2018), using a standard proficiency test, the specific study focuses more on acculturation (by solely testing a group of highly proficient Croatian- German immigrants) rather than purely the FLE. The authors investigated the effect of native-ness on moral judgment in decision making using personal moral dilemma scenarios. The authors hypothesized that the FLE might be diminished or absent in speakers with very high foreign language proficiency. What is more, they test whether high acculturation into the L2 culture (by immigrating) reduces utilitarianism in the L2. The findings of the study did not reveal a significant difference between the two languages, therefore no FLE as the participants' responses were the same in both Croatian and German. However, more utilitarian responses were recorded in the manipulated (more emotional) versions of the scenarios (by raising the number of people to be saved from 5 to 20) and adding an even more emotional sentence to give vividness to the scenario) in contrast to the original versions. Consistent with Harris (2004) and Costa et al. (2014) that found a proficiency effect; more deontological responses were recorded in participants with higher L2 proficiency; have also found a modulating role of emotion; high L2 proficiency increases emotionality, making proficiency effects in relation to the cognitive load hypothesis questionable. But once again the validity of the proficiency scores could be questioned as only multiple-choice questions were used to measure the FL Proficiency test, adapted from two different websites, instead of a standard FL proficiency test.

To sum up, the studies above collectively examine the FLE in moral decision-making, highlighting the role of language proficiency. Wong & Ng (2018) found that FLE diminishes when bilinguals are highly proficient in their L2, with language dominance influencing moral choices. More difficult personal moral dilemmas were easier to resolve in the L2, while emotional stimulation increased utilitarian responses regardless of language dominance. Similarly, Cavar & Tytus (2018) explored how high L2 proficiency and acculturation impact moral judgment, finding no significant FLE among highly proficient Croatian-German immigrants. Their results align with prior

studies suggesting that higher L2 proficiency enhances emotionality, complicating the cognitive load hypothesis. Both studies emphasize the complex interplay between proficiency, emotion, and moral decision-making, though questions remain regarding the accuracy of self-reported or non-standardized proficiency measures.

On the other side, there are some studies that did not seem to find a significant effect of proficiency on the FLE, raising the issues about the validity of the cognitive load hypothesis. For instance, Privitera et al., (2023), who measured FL proficiency using self-rated questionnaires highlight limited evidence indicating that FL proficiency and immersion impact decision-making with significant effects in the FL on the personal versions of the *organ transplant* and *burning building* dilemmas, respectively. The authors, however, report the exclusively self-rated proficiency as a possible limitation to their study (studies that used self-rated proficiency and did not find an effect of proficiency on the FLE: Dylman & Bjärtå, 2019; Geipel et al., 2015; Geipel et al., 2016; Hayakawa et al., 2017; Privitera et al., 2023; Vives et al., 2018) see also metanalyses by Circi et al., 2021; Del Maschio et al.,2022). The previous statement emphasizes the need for implementing more explicit/standard measures of FL proficiency within the framework of the FLE.

Miozzo et al. (2020) examined whether decision-making differences observed in foreign languages also appear in regional languages spoken fluently by native bilinguals. Prior research suggests that using an FL reduces cognitive biases and promotes utilitarian moral judgments (Costa et al., 2014; Keysar et al., 2012), often attributed to weakened emotional processing (Hayakawa et al., 2016). However, these effects had only been investigated in late-learned, school-taught languages, rather than in bilinguals who speak both languages from early childhood. To address this, the study examined Italian-Venetian and Italian-Bergamasque bilinguals, whose regional languages are acquired naturally and used primarily in informal settings, while Italian serves as the dominant language for formal, public, and institutional contexts. Four experiments tested whether decision-making in regional languages aligns with native or foreign languages, whether emotional processing differs across languages, and whether proficiency influences these effects. Study 1 measured emotional intensity ratings in Venetian and Italian using self-assessment manikin (SAM) scales and found no significant difference in emotional

responses claiming a challenging of the hypothesis that emotional detachment drives the foreign language effect. However, it is important to mention that emotional intensity was not measured in relation to moral judgment. Study 2 used the Asian Disease Problem to examine the framing effect, revealing that framing biases persisted in Italian but disappeared in Venetian, replicating prior findings in foreign languages (Costa et al., 2014). Study 3 tested moral decision-making in the *footbridge dilemma*, showing that utilitarian choices increased in Venetian compared to Italian, mirroring results from foreign language research (Geipel et al., 2015; Hayakawa & Keysar, 2018). Study 4 replicated these moral decision-making findings in Italian-Bergamasque bilinguals, reinforcing the conclusion that decision-making biases observed in foreign languages also appear in regional languages. To examine the role of proficiency, participants' Venetian and Bergamasque proficiency was measured using self-ratings (1–10 scale) and an objective language test, where they identified grammatical correctness in minimal pairs of Venetian/Bergamasque sentences. However, participants with low proficiency scores were excluded from the final analysis as only highly proficient bilinguals were considered, therefore, even though this study used objective proficiency tests for two of their studies there was no comparison between low and high proficient participants. What's interesting in their findings was that despite controlling for proficiency, no significant effects were found on moral judgment, indicating that decision-making differences were not influenced by fluency levels-challenging therefore the cognitive load account suggesting that the observed shifts toward reduced cognitive biases and increased utilitarian judgments in regional languages were not driven by language mastery, but by sociolinguistic factors, such as the contexts in which these languages are used (Miozzo et al., 2020). However, Miozzo et al. (2020) challenge the cognitive load explanation by showing that even highly proficient bilinguals, who speak their regional languages fluently and use them regularly, exhibit similar decision-making shifts to those found in foreign language studies. If cognitive load was the driving factor, one would expect no such differences between Italian and Venetian/Bergamasque, as participants were equally proficient in both languages. Since no proficiency effects were found, the study suggests that contextual and sociolinguistic factors, rather than cognitive effort, may better explain why decision-making shifts occur in different languages (Miozzo et al., 2020).

A meta-analysis paper on how language proficiency influences the FLE, based on existing research, emphasizes the difficulties of relying on subjective self-rating measuring for language proficiency as in spite of sufficient variability in the sampled data, no moderating effect of L2 proficiency on the FLE was found (Circi et al., 2021). The authors suggest for the use of objective measures of proficiency as primary studies relied on subjective self-reports, which also differed considerably across studies (Marian et al.,2007).

To sum up, the cognitive load hypothesis highlights that using a FL requires greater mental effort than a native language due to unfamiliar linguistic structures and vocabulary (Majerus, 2013; Majerus et al., 2019). This heightened cognitive demand often limits decision-making resources, leading to reliance on heuristics—mental shortcuts that can result in less optimal outcomes (Kahneman, 2011; Costa et al., 2014). Based on the abovementioned studies, FL proficiency plays a key role in reducing cognitive load, with higher proficiency enabling faster processing and fewer decision-making biases (Schweizer, Craik, & Bialystok, 2013; Kroll & Bialystok, 2013; Costa et al., 2014). However, research findings are variable, often due to reliance on self-reported proficiency, which can lack reliability (Tomoschuk et al., 2019). Studies emphasize the need for more standardized proficiency assessments to better understand the interplay between FL skills, cognitive load, and decision-making, providing a clearer framework for evaluating the FLE (Circi et al., 2021; Marian et al., 2007).

Table 1 summarizes 27 studies on the FLE. Among these, 19 studies focus on moral decision-making, including Białek et al. (2019), Brouwer (2019, 2021), Cavar & Tytus (2018), Chan et al. (2016), Cipolletti et al. (2016), Corey et al. (2017), Costa et al. (2014), Dylman & Champoux-Larsson (2020), Geipel et al. (2015), Hayakawa et al. (2017), Hayakawa & Keysar (2018), Kyriakou et al. (2023), Miozzo et al. (2020), Muda et al. (2018), Wong & Ng (2018), and Woumans et al. (2020).

While the majority of studies examine moral judgment as a specific domain, others explore related areas such as risk-taking decisions (Dylman & Champoux-Larsson, 2020; Hadjichristidis et al., 2015; Hadjichristidis et al., 2019), affective valence (Hadjichristidis et al., 2019; Vives et al., 2021), and emotion-based decisions

(García-Palacios et al., 2018; Zheng et al., 2020). In total, 57 experiments were included across these studies, offering a broad and detailed exploration of how the FLE impacts moral decision-making. This extensive body of work provides strong empirical evidence that using an FL influences various forms of judgment.

Table 1

Summary of FLE studies on moral decision making, risk taking and other emotional based decisions.

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
Costa, Foucart, Arnon, Aparici, & Apesteguia, et al.,(2014)	1	English, Spanish, Korean, French, Hebrew	317	Personal moral dilemma (between-Subjects)	Footbridge Dilemma	N/A (inferred from previous studies).	Self-rated	FLE more utilitarian choices in FL.	Moral Judgment	Moral decision-making
	2	Spanish English,	725	Personal and impersonal moral dilemmas (between subject).	Footbridge & Switch Dilemmas	N/A (inferred from personal/impersonal)	Self-rated	FLE only for personal dilemma more utilitarian choices in FL). Higher proficiency reduced FLE.	Moral Judgment	Moral decision-making
Geipel et al. (2015)	3	Korean English	161	Personal moral dilemma (between-Subjects)	Trolley, fumes, transplant, crying baby	N/A	Self-rated	FLE only for personal dilemmas Proficiency correlates with FLE	Moral Judgment	Moral decision-making
Cipolletti, McFarlane & Weissglass (2016)	4	English Spanish Spanish English	160	Moral dilemmas (between subjects)	Footbridge and trolley (personal/ impersonal)	NA	Self-rated	FLE only on personal dilemma.	Moral judgment	Moral decision-making
Chan, Gu, Ng, & Tse, (2016)	5	English Chinese	14	Moral dilemmas (between-subjects)	39 personal& impersonal moral dilemmas (e.g. footbridge, trolley, lifeboat, donation, environmental policy, vaccine...).	Self-rated	Self-rated	No FLE. Higher emotional arousal was linked to more utilitarian choices.	Moral judgment	Moral decision-making
	6	English Chinese	4	Moral dilemma (between-subjects)	Footbridge dilemma (personal).	Self-rated	Self-rated	FLE on footbridge dilemma but not mediated by emotional arousal.	Moral judgment	Moral decision-making
Corey et al., (2017)	7	English Spanish	211	Moral dilemmas (between-subjects)	Footbridge, Switch); Vocabulary translation task	NA		FLE on moral judgment. Higher vocabulary	Moral judgment	Moral decision-making

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
							Self-rated & vocabulary translation	knowledge did not eliminate the effect.		
	8	English Spanish	173	Moral dilemmas (between-subjects)	Moral dilemmas (Hospital, terrorist)	NA	Self-rated	FLE on the Terrorist dilemma but not in the Hospital dilemma.	Moral judgment	Moral decision-making
	9	English Spanish	204	Moral dilemmas (between-subjects)	Footbridge, Switch); Language switching task	NA	Self-rated	FLE only in personal dilemma.	Moral judgment	Moral decision-making
	10	English Spanish	399	Moral dilemmas (between-subjects)	Footbridge, Switch; Social group membership manipulation	NA	Self-rated	FLE on personal dilemma regardless of group membership (in-group/out-group).	Moral judgment	Moral decision-making
	11	English Spanish	202	Moral dilemmas (between-subjects)	Footbridge with button, Switch.	NA	Self-rated	FLE on the modified Footbridge. Action aversion had a small impact.	Moral judgment	Moral decision-making
	12	English Spanish	190	Moral dilemmas (between-subjects)	Footbridge, Switch - Reframed questions on inaction consequences	NA	Self-rated	FLE in Footbridge when inaction consequences were highlighted.	Moral judgment	Moral decision-making
	13	English Spanish	201	Moral dilemmas (between-subjects)	Moral dilemmas (Footbridge, Switch) - Explicit trade-off framing	NA	Self-rated	No FLE on footbridge when both action and inaction trade-offs were highlighted.	Moral judgment	Moral decision-making
	14	English Spanish	197	Moral dilemmas (between-subjects)	Moral dilemmas (Footbridge with disability consequence, Switch)	NA	Self-rated	FLE when the consequence was disability.	Moral judgment	Moral decision-making
	15	English Spanish	223	Moral dilemmas (between-subjects)	Moral dilemmas (Footbridge with injury consequence, Switch)	NA	Self-rated	No FLE when the consequence was injury.	Moral judgment	Moral decision-making
	16	English Spanish	211	Moral dilemmas (between-subjects)	Footbridge, Switch); Vocabulary translation task	NA	Self-rated & vocabulary translation	FLE on moral judgment. Higher vocabulary knowledge did not eliminate the effect.	Moral judgment	Moral decision-making
Hayakawa, Tannenbaum, Costa, Corey,	17	German English	214	Incongruent vs. congruent version of moral dilemmas	Time machine, car accident, hard times. Crying baby, relationship,	Process Dissociation Task	Self-rated	Clear FLE across all studies.		
	18	English Spanish	242							

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
& Keysar, (2017)	19	Spanish English	195	(between-subjects)	abortion, torture, vaccine policy, animal research, border crossing.	Congruent vs. Incongruent Moral Dilemmas		FL use blunts emotional reactions/deontological responses.	Moral judgment	Moral decision-making
	20	German English	211							
	21	German English	209							
	22	English German	206							
Cavar&Tytus, (2018)	23	Croatian German	60	More dilemmas (3 personal & 3 personal with intensified emotionality e.g. killing 20 instead of 5 people)	Footbridge, hostages, Submarine, soldiers, life raft, surgery	More emotional scenarios.	FL proficiency test	No FLE on moral dilemmas but more utilitarian responses in the more emotional versions of the scenarios.	Moral judgment	Moral decision-making
Hayakawa & Keysar (2018)	24	German English	800	Moral dilemma (personal) (between-subjects)	Footbridge dilemma (moral decision, mental imagery vividness)	Indirectly measured via mental imagery ratings.	Self-rated	FLE: using a foreign language reduces visualization of the victim, leading to more utilitarian moral choices.	Mental imagery moral judgement	Moral decision-making
Muda et al., (2018)	25	Polish English	140	Moral Dilemmas (between-subjects)	Incongruent vs. congruent version of moral dilemmas	NA	Self-rated	No FLE on moral judgments	Moral judgment	Moral decision-making
Wong & Ng (2018)	26	English Chinese	86	Personal and impersonal moral dilemmas (between subject).	Burning Building, Crying Baby, Organ Transplant, Shark Attack, and Trolley/Footbridge	Self-rated	Self-rated (early bilinguals)	FLE only on personal dilemmas. No Strong FLE in early bilinguals. Emotional arousal increased utilitarian choices but was unrelated to language dominance	Moral judgment	Moral decision-making
Białek, Paruzel-Czachura, & Gawronski, 2019	27	Polish English	204	Moral Dilemmas (between-subjects)	Proscriptive norm (forbids action) vs. prescriptive norm (requires action). High vs. low benefit-cost ratio for the action. (Personal/Impersonal e.g hospital virus outbreak, hostage	NA	Self-rated	No FLE on moral judgments but low sensitivity to consequences and norms in the FL.	Moral Judgment	Moral decision-making
	28	Polish German,	138							
	29	Polish Spanish	163							
	30	Polish French	129							

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
					negotiation, train switch, poisoned well, airplane highjack etc)					
Brouwer, (2019)	31	Dutch English	60	Moral Dilemmas (between-subjects)	Text (reading): Personal& Impersonal dilemmas (footbridge, crying baby, vitamins, trolley, lost wallet, taxes)	NA	Self-rated	No FLE on moral judgments Reading allows deliberation, possibly reducing the FLE	Moral Judgment	Moral decision-making
	32		60		Audio (listening): listened to recorded versions of the same six moral dilemmas			FLE on moral judgement on the auditory condition on both personal/impersonal.	Moral Judgment	Moral decision-making
Dylman & Champoux-Larsson, (2020)	33	Swedish-English	198	Moral dilemma (between-subjects)	Footbridge Dilemma	NA	Self-rated	No FLE Culturally influential	Moral Judgment	Moral decision-making
	34	Swedish-French	175					FLE Less culturally influential		
	35	Swedish-Norwegian	305	Moral dilemma (between-subjects) linguistic Similarity)				No FLE - Linguistically similar		
	36	Norwegian-Swedish	295							
Miozzo et al. (2020)	37	Italian Venetian	408	Phrase Rating Task Anchor Contraction Effect - ACE	Listened to emotional phrases (e.g., endearments, insults, reprimands)	Rated how much emotion (e.g., fear, sadness, happiness) Self-Assessment Manikin (SAM)	Self-rated Language proficiency test to verify fluency. 7-item grammaticality	No difference in emotionality between languages. No role of proficiency on the FLE.between high and low groups.	Moral Judgment Moral Judgment	Moral decision-making
	38	Italian Venetian		Moral dilemma (between-subjects)	Asian Disease problem		Self-rated	No framing effect in Venetian, unlike Italian	Moral Judgment	Moral decision-making
	39	Italian Venetian	195	Moral dilemma (between-subjects)	Footbridge dilemma	Self-rated	NA	FLE	Moral Judgment	Moral decision-making

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
	40	Italian Bergamasque		Moral dilemma (between-subjects)	Footbridge dilemma	Self-rated Language proficiency test to verify fluency. 10-item grammaticality test	NA	FLE No role of proficiency on the FLE.	Moral Judgment	Moral decision-making
Woumans et al. (2020)	41	Dutch English	558	Personal an impersonal crime scenarios (between-Subjects	Murder crime scenarios (personal and impersonal).	N/A	Self-rated	FLE on all crime-related moral dilemmas. FL diminishes crime severity judgment No effect of proficiency.	Moral judgment	Moral decision-making
Brouwer (2021)	42	Dutch English	154	Personal and impersonal moral dilemmas (between subject).	Trolley and footbridge dilemmas (visual and auditory modalities)	N/A	Self-rated	FLE on personal dilemmas. FLE not dependent on modality. FLE on high proficiency for personal dilemmas. Emotional attenuation was stronger in auditory FL contexts.	Moral judgment	Moral decision-making
Kyriakou, Foucart, and Mavrou (2023)	43	Spanish English	204	Personal Dilemma	Footbridge dilemma	Open-ended questions to explain reasoning & emotion	Self-rated	FLE mediated by emotions. L1 responses contained significantly more high-arousal words, while L2 responses included more low-arousal words.	Moral judgement	Moral decision-making
Geipel et al. (2015)	44	German English Italian English	38 60	Scenarios (between-subjects)	Dog, incest, exam, and flag scenarios	N/A	N/A	FLE (lower moral wrongness in FL)	Moral transgression	Moral decision-making
	45	Italian, English	78	Scenarios (between-subjects)	Dog, incest, exam, and flag scenarios	Self-rated	N/A	FLE on high-emotion violations (dog, incest) but not for low-emotion violations (exam, flag)		
	46	Italian, English German Italian	74	Scenarios (between-subjects)	Dog, incest, exam, and bonus scenarios confidence in moral judgment Moses illusion task	Self-rated	Self-rated	FLE on moral judgments, reduced confidence, and no improvement in analytic reasoning (Moses illusion task performance was worse in foreign language).		

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
Dylman & Champoux-Larsson, 2020	47	Swedish English	663	Moral dilemma (between-subjects)	Asian Disease Problem	NA	Self-rated	No FLE Culturally influential	Risk aversion	Risk taking decisions
Hadjichristidis, Geipel, & Savadori, 2015	48	Italian English	92	Affect in Judgments of Risk and Benefit tasks (between-subjects)	Specific hazards such as "traveling by airplane," "climate change," and "biotechnology.	NA	Self-rated	FLE on risk judgement.	Risk judgement	Risk taking decisions
Hadjichristidis, Geipel, & Surian (2019)	49	Italian English	123	Affect in Judgments of Risk and Benefit tasks(between-subjects)	Specific hazards such as "traveling by airplane," "climate change," and "biotechnology.	Self-rated Positive/ negative	Self-rated	FLE on risk judgement. Emotional reduction on negative emotions on judgment of risk and intensified positive ones on benefit in FL	Risk judgement	Risk taking decisions
Dylman & Bjärtå (2019)	50	Swedish English	34	Read and answered questions about negative and neutral texts in L1 and L2.	Texts in L1-L1 vs texts in L1-L2	Self-rated distress levels (before or after the questions)	Self-rated	Decrease of distress L2 use can diminish levels of distress experienced following a negative event encoded in one's first language.	Psychological distance	Emotional based decisions
Shin & Kim (2017)	51	Korean English	26	Self-bias effects	Matching tasks 'self' compared to the conditions 'friends'or 'others'	N/A	Self-rated (proficiency+ fluency)	Reduced self-bias in FL. Reaction times in matching tasks were reduced in the condition 'friends 'or 'others'.	Psychological distance	Emotional based decisions
Vives, Aparici, & Costa (2021)	52	Spanish English	26	Affect Labeling, Gender Labeling, Amygdala Activation	fMRI scan while labelling emotional faces either in their native or foreign languages.	fMRI scan	Self-rated (who had B1+ Cambridge English Qualifications)	No FLE on emotional arousal not downregulate emotions;	Affect labelling	Emotional based decisions
Hadjichristidis, Geipel, & Surian (2019)	53	Italian NL German FL English FL	181	Superstitious scenarios: Bad luck scenarios	Ladder Mirror	Self-rated	Self-rated & Translate scenarios	Reading information in a foreign language can suppress common superstitious beliefs.	Affective valence	Emotional based decisions
	54	German English	142	Superstitious scenarios: Bad luck & good-luck scenarios (between-subjects).	Bad luck: Friday 13th mirror, black cat. Good luck: Four-leaf clover, falling star. Magical beliefs scale	Self-rated	Self-rated	FLE: bad-luck and good-luck superstitions. Less negative feelings towards bad-luck scenarios and less positive feelings towards good-luck scenarios (no influence on non-superstitious, control	Affective valence.	Emotional based decisions

Study		Languages	N	Design*	Tests	Emotionality	Proficiency	Results	Domain specific	Domain general
								scenarios). No FLE on magical beliefs.		
	55	German English	435	Superstitious: bad luck & good-luck scenarios vs control scenarios. (between & within subjects)	Same bad luck & good luck scenarios. Control scenarios (blocked sink, white dog, tulips, airplane) Magical beliefs scale.	Self-rated	Self-rated	FLE: attenuated both bad-luck and good-luck superstitions. No FLE or emotionality effect on magical beliefs when comparing results from study 2+3.	Affective valence.	Emotional based decisions
Zheng, Zhang, Wang, & Li (2020)	56	Turkish English	32	Compared emotional and physiological responses to emotionally charged words in L1 vs. L2.	Electrodermal monitoring, word recognition (taboo, reprimand, neutral, positive, aversive)	Electrodermal monitoring	NA	FLE: Greater skin conductance responses to taboo and childhood reprimand words in L1 than L2; auditory presentation provoked stronger responses than visual.	Emotional responsiveness to taboo words	Emotional based decisions
García-Palacios et al., (2018)	57	Spanish English	54	Cue a potential threat (CS+) Indicate safety (CS−)	Verbal instructions (NL/FL). Counting down aloud from 10 to 1 while being presented with these stimuli. Coloured geometric shapes	Measured through pupil dilation and skin conductance responses (emotional reactivity).	Self-rated	Fear conditioning was more effective in the NL. FL detachment effect (Fear response was reduced in the FL).	Fear conditioning	Emotional based decisions
Hayakawa & Keysar (2018)	58	German English	800	Moral dilemma (personal) (between-subjects)	Footbridge dilemma (moral decision, mental imagery vividness)	Indirectly measured via mental imagery ratings.	Self-rated	FLE: using a foreign language reduces visualization of the victim, leading to more utilitarian moral choices.	Mental imagery and moral judgement	Emotional based decisions

Note. Between-participant and within-participant design in relation to the language tested (different groups different languages, or same group tested in both languages).

This PhD research aims to offer a significant and novel contribution to the existing literature on the FLE by systematically investigating the role of FL proficiency

in decision-making. Notably to Miozzo et al. (2020), who exclusively examined highly proficient bilinguals and found no effects of proficiency on FLE, this study will compare both low- and high-proficiency participants to determine whether proficiency moderates decision-making biases. By incorporating participants with varying levels of language proficiency, this research directly tests the cognitive load hypothesis on the FLE. If low-proficiency bilinguals exhibit a stronger FLE—characterized by increased utilitarianism while high-proficiency bilinguals show weaker or no effects, this would lend empirical support to the cognitive load hypothesis. This hypothesis posits that increased cognitive effort when processing a foreign language reduces reliance on intuitive, affect-driven reasoning, thereby promoting more analytical decision-making. Conversely, if both proficiency groups exhibit similar effects, this would reinforce the conclusions of Miozzo et al. (2020), who argued that the FLE does not arise due to cognitive resource demands.

2.3.2 Foreign Language Fluency on the FLE

Interconnected with FL proficiency another aspect of the cognitive load hypothesis is the FL fluency. FL fluency is the capacity to communicate effectively in a language other than one's native tongue, encompassing speaking, reading, writing, and comprehension skills. This measure includes both linguistic competence, such as vocabulary and grammar, and pragmatic competence, which is the ability to use language appropriately in various social contexts. Achieving fluency means being able to engage in conversations and formal interactions naturally and effectively (Council of Europe, 2001; ACTFL, 2012; Canale & Swain, 1980; Krashen, 1982; Bachman & Palmer, 1996).

It is important to note that even though language proficiency and fluency are interconnected are yet distinct concepts, with proficiency encompassing a broad range of language abilities such as comprehension, vocabulary, and grammar, while fluency emphasizes the smoothness and spontaneity of language use, particularly in speech. Proficiency provides the foundation for fluency, as grammatical and lexical knowledge enable fluid and effective communication. However, fluency captures a dynamic, real-time processing skill that goes beyond mere accuracy, often reflecting conversational

ease and adaptive language use (Bakkouche & Saito, 2025; Yan et al., 2025). As such, measuring fluency provides unique insights into communicative competence that proficiency assessments may not fully capture (Shekarabi & Ebrahimi, 2025; Haake, 2025).

According to Hayakawa et al. (2016), encountering disfluency in an FL might significantly contribute to the FLE. This is because the challenge of processing disfluency can prompt a more thoughtful and meticulous approach to thinking, as the above-mentioned cognitive load hypothesis. Higher fluency facilitates more automatic and less effortful language processing, potentially reducing the deliberative thinking typically associated with the FLE (Hayakawa et al., 2016). Conversely, lower fluency requires more cognitive resources, leading to more deliberate and reflective thinking (Hayakawa et al., 2016; Council of Europe, 2001; ACTFL, 2012). This heightened cognitive load can amplify the FLE, as individuals engage in more careful analysis and decision-making processes when operating in a less familiar language. As an additional measure of this thorough investigation of the role of the cognitive load hypothesis language fluency will be examined in experiment 1.

To sum up, this PhD investigates the role of FL proficiency and fluency; whether low FL proficiency leads to more cognitive load; and fluency on the FLE in experiment 1. It addresses this by objectively measuring proficiency and fluency scores (from standard tests) in low and high proficient bilinguals of each dilemma that should provide evidence no other study has done before. Based on the theory review of this chapter, on the existing number of studies on the two most distinct parameters that could drive the FLE (see Hayakawa et.al.,2016) it seems that it is still unclear on whether the FLE is driven by emotionality or cognitive load processing in the FL.

2.4 Methodological issues

2.4.1 Self-preservation and non-self-preservation moral dilemmas

The FLE leads to more utilitarian decision-making in personal dilemmas by reducing emotional responses, whereas impersonal dilemmas, which are less emotionally intense, do not show a significant FLE (Corey et al., 2017; Costa et al., 2014; Hayakawa et al., 2017).

The majority of previous research that used moral dilemmas did not acknowledge a possibly important difference between these dilemmas, which is whether the decision (of killing someone to save 5-10 people) involved saving themselves along with others (Białek & Fugelsang, 2019). Previous research has revealed that it is morally acceptable to harm someone in self-defense. Making moral choices of killing one person to save others (utilitarian decisions) are more probable when the life of the person making the decision is in danger too in contrast to decisions that only involve other people's lives (Suessenbach & Moore 2015).

Mills & Nicoladis (2020) conducted a study that focused explicitly on two dilemmas; namely in this study; a self-preservation and a non-self-preservation dilemma, as aforementioned. A dilemma where a participant hypothetically kills one person to save the lives of many others and a dilemma where the hypothetical action of killing a person will concern saving one's own life along with others. As a non-self-preservation dilemma, they used the *footbridge dilemma* that involves pushing a bulky man who is standing on a bridge to fall in front of a trolley, stop its course and save five people. As a self-preservation dilemma they used the *crying baby dilemma*; where the participant and their neighbours are hiding from enemy soldiers in a war situation, their baby starts crying and will attract soldiers that will kill them all unless the participant puts their hand on their baby's mouth to make it stop crying, but as a result the baby dies and everyone else is saved. The results of the study revealed that 'It's easier to kill a baby to save oneself than a bulky man to save other people' providing evidence that

even though these are both personal dilemmas the type of dilemma (SP or NSP) plays an important role in either language (Chan et al. 2016).

A recent study by Privitera et al., (2023) classified moral dilemmas from Christensen et al. (2014) (*organ transplant, burning building* and *footbridge dilemma*; 3 of the same dilemmas used in the current study) not only in terms of personal and impersonal versions (see Wong & Ng , 2018) but also in Self-beneficial dilemmas; entailing scenarios where the decision-maker's own life is in danger vs Other-beneficial dilemmas; solely involve the well-being of others which are referred to in the current PhD thesis as self-preservation and non-self-preservation moral dilemmas). Participants had to make a "yes" or "no" decision on whether they would commit the action described in the moral dilemma and were then directed to a new screen where they were prompted to assess the acceptability of the action described in the dilemma using a 7-point Likert scale. Finally, participants were prompted to evaluate their understanding of the dilemma on a scale from 1 to 7. Self-rated proficiency was used, and the results revealed that the FLE varied across dilemmas, with significant effects observed on some, like the Organ Transplant dilemma, but not others. The results revealed a reverse Moral FLE which means that participants showed a higher tendency to provide a utilitarian response when they encountered the *burning building dilemma* (self-preservation dilemma) in their native language, where participants were less inclined to commit a moral violation compared to the *footbridge dilemma* where no such effect was found. This contradicts previous findings by Wong & Ng (2018) which didn't show significant results. The *burning building dilemma* is the only scenario where participants' own survival is at risk if they don't commit the moral violation in Privitera et al., (2023) study. The authors explain that a possible explanation for this could be that participants are more sensitive to negative emotions when faced with dilemmas that directly impact their own survival. Additionally, these self-beneficial dilemmas might be perceived as more emotionally charged compared to others.

2.4.2 Within and between subjects designs and induced Language mode

Wong, & Ng (2018) argue that the FLE on moral judgement decisions has so far been tested mostly by comparing results from different participants. Previous studies followed the 'in between subjects' comparisons by dividing participants into two separate groups collecting data in either the native or the FL (e.g., Cipoletti et al., 2016; Cushman, 2013; Geipel et al., 2015; Keysar et al., 2012). However, this type of comparison can introduce some undesirable effects as the responses compared derive from different individuals with different L2 background (language proficiency, age of acquisition, L2 immersion etc.). The current research will attempt to collect data using a 'within groups design' (e.g., Reyes, 2020) comparisons in both languages, from the same participants.

The attempt will be controlled by manipulating language mode between tasks to some extend on the target population (Greek native speakers). Language mode refers to the level of activation of a bilingual individual's languages and the cognitive mechanisms involved in language processing at a specific moment (Grosjean, 2008). The short video (neutral topic will manipulate the language mode of the language that will follow eliminating L1 influence on the L2 and vice versa on the FLE (Elston-Güttler et al., 2005). The short videos were supplemented by comprehension questions in the target language to thoroughly involve participants in the task. Responses with no answers to the comprehension questions were discarded.

Additionally, since this was an online questionnaire with written instructions only in the target language it adds environmental language control to some extent as no oral instructions will be given by a bilingual Green & Abutalebi (2013). This opportunity will help us weigh evidence towards the role of cognitive load (L2 proficiency) or the reduced L2 emotionality hypothesis in relation to the FLE. The standard proficiency level representation (Oxford QPT) will provide solid results that could be then related to the FLE.

However, it is important to mention that according to Athanasopoulos (2016), apart from language proficiency there are other factors relating to frequency of language use and amount of immersion in the L2 context. Grosjean (1998), discusses several

factors that may underpin linguistic and cognitive processing in the bilingual person. Factors such as the frequency of L2 use and amount of immersion in a L2 context (e.g., to live in an L2 speaking country (acculturation) can shape both linguistic and non-linguistic way of acting. This information was collected using a language background questionnaire in order to test some of these factors that might influence the FLE in an exploratory data analysis.

2.5 Rationale of the study

2.5.1 Related research

Costa et al, (2014) reported findings that our judgement on moral dilemma changes in a foreign language. The authors make the hypothesis that morally judging a dilemma in a language other than your native will result in reduced emotional reactivity. They predict that utilitarian judgments (judgments that favour the greater good despite an individual's rights) will occur more in a foreign language (FL) in contrast with L1. Two experiments were carried out in order to test the above hypothesis. In the first scenario the authors used the *footbridge dilemma* where participants have to picture themselves standing on a footbridge above a train track, where they notice a train heading toward five people. The only way to stop it (to avoid killing them) is by pushing a heavy man off the bridge onto the tracks, sacrificing him to save the five.

The authors' prediction implies that when the dilemma is presented in a FL, there will be more chances that participants will sacrifice one man to save five (i.e., utilitarian response), whereas this won't be common in their L1 as it is morally and emotionally upsetting since in involves the action of killing one man. The results of the first experiment substantiate the hypothesis revealing that more than half of the responses in the FL were utilitarian decisions; due to lessened emotional significance; in contrast to the L1. In the second experiment, however, the authors used the *trolley dilemma* which is less emotionally unpleasant since the participants have to hypothetically pull a switch, to change the trail, killing only one person instead of five and not physically push someone like in the *footbridge dilemma*. The results revealed fewer utilitarian choices in the L1 than in the L2 (regardless of the type of dilemma).

Yet, the poor proficiency measures employed (self-rating proficiency scores; Tomoschuk et al., 2019) and the between subjects design create a possible doubt onto whether the responses in the footbridge condition are due to L2 proficiency, and not down to some unknown variable that may be confounding results across the two groups of participants due to the between-subjects design (Białek and Fugelsang, 2019).

Similarly, Wong, & Ng (2018) used ten scenarios originally brought together in a study by Christensen et al. (2014); asking participants to state whether they would commit a certain utilitarian action, killing one person in order to save five to eleven people. The scenarios included both personal and impersonal versions of the dilemmas; In personal moral dilemmas, the individual is directly causing harm, whereas in impersonal moral dilemmas, the individual's involvement is indirect and part of a process that leads to harm (Christensen et al. 2014); revealing that personal dilemmas are more likely to lead to a personal decision involving the use of force, as opposed to impersonal ones (Wong & Ng's, 2018), which was revealed in the results of the specific study where there was no significant FLE on impersonal dilemmas. Therefore, since proficiency may affect responses exclusively to personal moral dilemmas, the current study focuses in on this type of dilemmas, predicting a weaker FLE in higher proficiency FL learners than low proficiency FL learners.

2.5.2The current study

As reviewed in this thesis several previous studies have investigated the FLE on moral judgment in various foreign languages. However, the need for further study in more languages and in more methods is stressed out by Hayakawa et.al, (2017). To date, no study has specifically examined Greek native speakers in this exact manner; thus, this research contributes to the existing body of literature.

One of the aspects that is not clear in FLE research is whether the FLE is a result of lower proficiency and fluency in the FL as previous research has used poor FL proficiency measures and no fluency measures in general. Previous studies mostly tested language proficiency through participant self-reports (Costa et al., 2014; Hayakawa & Keysar, 2018; Wong & Ng, 2018). This will be one of the first studies to examine the FLE on moral judgment by providing a standard language proficiency test

providing each participant's formal language skills score complying with the international CEFR scale.

The first major study asked participants to morally judge some harmful hypothetical personal moral dilemmas which require participants to make the utilitarian decision of killing/ sacrificing a person to save more (1-10 people) (Wong & Ng, 2018). The dilemmas were divided into two groups; SP and NSP something that was not acknowledged in previous similar research (Cavar & Tytus, 2018; Wong & Ng, 2018). First, the role of language proficiency was investigated (using a solid proficiency level measure) and then the role of foreign language fluency was examined (using standard fluency measures) on the FLE. Additionally, the role of emotional reduction in the FL was elicited, in a first attempt, by asking participants to rate the perceived emotionality of each dilemma) in an attempt to weigh evidence towards the cognitive load vs language emotionality, or both.

The second experiment investigated exclusively the role of emotions on the FLE on moral judgement by employing rigorous emotion rating measures in an attempt to investigate the emotional states prior and after participants had to make a hypothetical moral decision and the role of FL use respectively.

2.6 Research questions

2.6.1 Experiment 1

The aim of the current study was to attain further insights on the role of foreign language proficiency and fluency on the FLE on moral judgement in more depth; in order to contribute to the literature gap proposed by Hayakawa's et. al. (2016) and test the theory that high language proficiency minimizes the FLE to some extent (e.g., Wong & Ng, 2018). To test the FLE on moral judgements I examined whether hypothetical harmful moral dilemmas are judged in a more utilitarian manner when using a foreign language instead of the native language. Emotionality ratings for each of the moral dilemma scenarios were also used. Furthermore, the specific group of participants; Greek-English bilinguals; weren't tested in existing literature so far along

with rigorous proficiency measures and acknowledgement of possible difference between SP and NSP moral dilemmas.

Based on previous research I predicted that the presentation of Moral judgements in a foreign language will lead to more utilitarian response (e.g., take one life to save five people) than in the L1, where in turn responses are expected to conform to moral imperatives (don't take a life). For instance, results presented in Circi et al., (2021) meta-analysis show that bilinguals tend to judge moral decisions less rationally in their L1 and /or less emotionally in the FL (Geipel et al., 2015, Hayakawa et al., 2016, Costa et al., 2014). Furthermore, Circi et al., (2021) metanalyses also compared the impact of language proficiency on the FLE. However, almost in all of the studies the language proficiency results derived from participants' self-perceived level of proficiency. Therefore, the importance of using a standardized method to measure FL proficiency, during the time FLE effect is measured, is becoming even more urgent in order to provide concrete results on the role of FL proficiency on the FLE (Tomoschuk et al., 2019).

One of the great unknowns in the FLE research is whether the effect is rooted in more processing effort resulting from lower proficiency and poorer fluency in the L2. Yet, at the same time FL proficiency has been poorly measured, and fluency has not been measured at all. This leads to the novelty of the current study where instead of using only self-rating proficiency and fluency questionnaires, used in previous studies, this will be the first study to investigate the FLE on moral judgment by providing a concrete representation on participants' current foreign language proficiency and fluency level by adding both:

- a standard language proficiency test in the FL
- fluency standard tests

For the purpose of measuring FL proficiency the standard Oxford University English Language placement test will be used. This quick placement test is designed to measure individuals' English language skills and provide results that are designed to correspond to the international CEFR scale (Oxford University Press, n.d..). Standard placement tests provide rigorous representations and are a reliable measure that could

be used across research experiments providing objective real time proficiency representation (Tomoschuk et al., 2019).

Notably, standard semantic and phonemic standard fluency tests like letter and category fluency tests (which were used in the current study) are fundamental tools that offer crucial insights into lexical retrieval abilities and offer a more comprehensive view of participants' foreign language abilities by engaging distinct cognitive processes and retrieval strategies (Henry & Crawford, 2004; Shao et al., 2014). These tasks require participants to produce as many words as possible that either start with a specific letter (letter fluency) or belong to a particular category (category fluency) within a set time frame, typically 60 seconds. Research highlights that letter fluency tests primarily engage vocabulary access and the ability to retrieve words from one's mental vocabulary, while category fluency tests rely heavily on semantic memory (Shao et al., 2014). Consequently, letter and category fluency tests deliver a comprehensive assessment of language fluency.

What is more, another important aspect is that previous research, mentioned in the meta-analysis by Circi et al., (2021) used the dual effect pattern where two groups of different participants were assigned to either the native or the FL condition (known as the dual-effect paradigm) and results from different participants were compared. For this reason, the authors suggest that moral decision-making should be examined on the same participants. The current study has created a within-subjects design, where the FLE will be tested on the same participants in an aim to provide a more accurate representation of the possible FLE by eliminating the uncertainty of results deriving by comparing data between people with different dispositional perspectives and qualities (Hu & Reiterer, 2009). In order to achieve this, a language mode task is used to induce the target language, along with some comprehension questions to fully engage participants in the language mode task.

As experiments 1 will focus primarily on investigating the role of that cognitive load hypothesis on the FLE the following research questions were addressed:

What is the role of foreign language proficiency on the FLE on moral judgement?

Based on previous research (Costa et al., 2014; Geipel et al., 2015). I predicted that the FLE on moral judgement, will be reduced or eliminated in the FL condition in more proficient bilinguals in contrast to less proficient bilinguals. According to the cognitive load hypothesis, the more proficient an individual is the more deontological their responses will be when having to make moral judgments, as less cognitive effort is required due to rapid processing. Whereas the less proficient an individual is the more cognitive effort will be required (Costa et al., 2014; Foucart et al., 2016; Geipel et al., 2015; Sweller, 1988). It is suggested that this foreign language comprehension takes the full attention from the decision-making process as individuals need to spend more time in the process of understanding the information (Anton et al., 2020).

What is the role of foreign language fluency on the FLE on moral judgement?

Testing foreign language fluency, in addition to proficiency, is crucial for a comprehensive evaluation of language skills. Fluency emphasizes the ability to use the language naturally and spontaneously, reflecting real-world communication scenarios. According to Hulstijn (2011), proficiency tests often focus on assessing knowledge of vocabulary, grammar, and reading comprehension, which, while important, do not fully capture a person's ability to communicate effectively in everyday situations. Fluency assessments, on the other hand, gauge how well individuals can maintain conversations, respond promptly, and express ideas without undue hesitation (Hulstijn, 2011). This distinction is particularly important in professional and academic settings where practical communication skills are essential. Research by de Jong et al. (2015) supports this view, suggesting that fluency testing can reveal discrepancies between theoretical knowledge and practical application, ensuring that individuals are truly capable of operating in a foreign language environment. Therefore, incorporating fluency assessments alongside proficiency tests provides a more holistic understanding of an individual's language capabilities, ultimately leading to more effective communication and better outcomes in multicultural contexts (de Jong et al., 2015). Hence similarly to FL proficiency it would be expected that lower fluency will also result in higher FLE as suggested in the cognitive load hypothesis as it would be expected that a highly fluent participant who is able to retrieve more words in the standard fluency tests (Henry & Crawford, 2004; Shao et al., 2014), for example the letter and category tests used in

experiment 1, in the available 60 seconds time limit requires less effort than an individual with lower fluency that will retrieve less words due to the limited vocabulary learnt.

What is the role of the type of dilemma SP/ NSP on the FLE?

Mills & Nicoladis (2020) state that people are more willing to commit the action of killing someone when it involves saving themselves too. The role of the type of dilemma, SP and NSP, will be further examined in the current study, however, more moral dilemmas will be used in contrast to the aforementioned study (that used only two). Also, a bigger sample size will be employed (107 instead of 71) along with even more rigorous proficiency measures (than the picture-naming task that was used in that study) and a within-subjects design alike to Mills & Nicoladis's study which will provide a more valid representation excluding external barriers that could arise in a between-subjects design that compares different groups of people (Białek & Fugelsang, 2019).

Secondary analysis will also explore what other variables may underpin the phenomenon of the FLE bearing in mind that it is still in its infancy.-For instance, age of acquisition may play an important role as the FLE is not found in bilinguals who learn the FL at a very young age (Brouwer, 2019; Cavar & Tytus, 2017; Wong & Ng, 2018), possibly because high linguistic capability and exposure to the language from a young age contribute in very similar emotional influence between the FL and L1. Furthermore, the length of stay in the L2 speaking country may also play a role as individuals are constantly exposed to the FL language in different contexts from everyday tasks to more complicated situations (Cavar & Tytus, 2017). Hence, other bilingual background variables that might affect the FLE on moral judgement (e.g., Athanasopoulos, 2016) were collected.

2.6.2 Experiment 2

The aim of this study was to focus solely in attaining further insights on the role of foreign language emotionality on the FLE on moral judgement in more depth; Once

again in an aim to investigate the literature gap on what drives the FLE proposed by Hayakawa's et. al. (2016).

There seems to be a FLE on moral dilemma judgement tasks hence moral dilemmas in the L2 are expected to produce more utilitarian responses (Cavar & Tytus, 2017; Wong & Ng, 2018). This could be due to the fact that emotionality is stronger in our L1 as it is developed from a young age simultaneously with morality (Holleman et.al,2021). Therefore, it is hypothesized that participants' emotional state will be affected more by moral dilemmas in the L1 rather than the FL. Hence, experiment 2 investigates the degree in which moral dilemmas induce more emotional responses in the L1 rather than the FL where based on the "dual-process theory of moral judgment" (Greene et al., 2001;Greene et al., 2008) FL decision making responses are less affected by emotions (Keysar et al., 2012). Duplicating the same design for the second experiment the presence of the FLE was once again tested with a new sample of Greek-English bilinguals. Here, I also predicted that Greek-English bilinguals will be more willing to commit the action of e.g. killing someone to save five others in their FL than in their L1 (Chan et al., 2016; Cipolletti et al., 2016; Costa et al., 2014; Geipel et al., 2015; Hayakawa et al., 2017; Wong & Ng, 2018). Such as, bilinguals will tend to judge moral decisions less rationally in their L1 and /or less emotionally in the FL (Circi et al., (2021).

- **What is the role of foreign language emotionality on the FLE on moral judgement?**

According to the "dual-process theory of moral judgment" there is a competition between the instinctive emotional responses and more thoughtful cognitive controlled responses. (Greene et al., 2001; Greene et al., 2008). An example of this could be moral dilemmas that regard physically harming someone (which instinctively activates emotional responses that predominate over the greater good) in contrast to dilemmas where emotion is not prominent resulting in controlled, utilitarian responses (Koenigs et al., 2007). The current study investigates the FLE on moral judgment by providing a concrete representation on participants' current emotional state employing an instrument that will provide us with the emotional states of participants prior to the

dilemmas and after the dilemmas in each language. The PANAS-X emotion rating (pre-to-post as found in Horne & Powell, 2016):

- Pre-test emotion ratings PANAS-X
- Post-test PANAS-X

Replicating the previous experiment design, we used a within-subjects design for the current experiment by testing the same participants in both languages in an aim to provide a more accurate representation of the possible FLE by eliminating the uncertainty of results deriving by comparing data between people with different dispositional

Chapter 3. General Methodology

3.1 Moral Dilemmas

The current study measured moral judgements by using four hypothetical personal dilemmas; dilemmas that require the agent to directly produce harm; which were selected from Wong & Ng (2018). The impersonal versions of the dilemmas in the original study were discarded as impersonal dilemmas showed no significant FLE on moral dilemmas judgement. Personal dilemmas are more likely to lead individuals to opt for a direct use of force, conversely to their impersonal versions.

Duplicating Wong & Ng (2018) methodological steps, participants will read the dilemmas and then the probable choice and impact before choosing (from a scale of 1 to7) how likely it would be that they commit the action in the given dilemma . In addition, participants will have to rate how upsetting each scenario felt (1 = most emotional, 7 = least emotional) in all experiments. Participants are hypothetically asked to make a utilitarian decision that will result in killing one person in order to save five others.

Each hypothetical moral dilemma indicated an action:

1. *Burning Building* (BB): Pushing an injured person to clear the debris of burning building in order to escape along with five other people,
2. *Footbridge* (FB): Pushing a large stranger onto the tracks to stop the train and prevent it from killing five workmen,
3. *Organ Transplant* (OT): Administer a higher dose of anaesthetic injection that will kill a critically injured patient in order to use the organs to save five other patients,
4. *Shark attack* (SA): Shoot at an injured diver in order for sharks to stop to eat him and give you and the other nine divers time to escape the sharks and be saved.

Moral dilemmas one and four are SP (*burning building, shark attack*) as participants will hypothetically save themselves along with others if they choose to commit the action. However,

moral dilemmas two and three (*footbridge, organ transplant*) are NSP dilemmas as they do not involve saving oneself.

The dilemmas were counterbalanced into four versions in all experiments: Version 1: English OT, Greek BB, Version 2: English: BB, Greek: OT, Version 3: Greek SA, English FB, Version 4: Greek: FB, English SA.

All items in Greek were translated and back translated by professional native Greek, English language teachers in order to ensure that both Greek and English parts were consistent.

3.2 Language mode in within-participant designs

Furthermore, another novel aspect that will be addressed in this study is that previous research (see Circi et al., 2021 & Białek & Fugelsang, 2019 for reviews) used between-subjects designs where two groups of different participants were assigned to either the native or the FL condition. The problem with the between-subjects design is that it attributes possible differences between the participants tested (i.e. ethical principles, cognitive approach, and memory capacity and working memory differ) whereas a within-subjects design will increase power as it mainly focuses explicitly on variances of the conditions and the experimental control (Białek & Fugelsang, 2019). Previous research has used multiple different dilemmas within the same participants in a single language (between-subjects design) and found no significant carryover effects on decision-making. Similarly, in my study, each participant is presented with each dilemma only once and never sees the same dilemma twice, ensuring that responses are not influenced by prior exposure to the same dilemma, similar to research by e.g., Costa et al., 2014; Hayakawa et al., 2017; Wong & Ng (2018)). Therefore, as I am using the same experimental approach where no participant is exposed to the same dilemma again carry over effects specific to each language context are not expected to arise. I have carefully selected several different dilemmas so that participants can make moral decisions in both languages, but for different dilemmas, allowing for a within-subjects design use while reducing potential biases.

The current study will implement a within-subjects design, where the FLE will be tested on the same participants, in both languages in all experiments, in an aim to provide a more accurate representation of the possible FLE by eliminating the uncertainty of results deriving by

comparing data between people with different dispositional perspectives and qualities (Hu & Reiterer, 2009).

In order to achieve this, a language mode (Grosjean, 1998) task was used to induce the target language, along with some comprehension questions to fully engage participants in the language mode task. Language mode refers to the degree of activation of a bilingual individual's languages and language processing systems at a specific moment (Grosjean, 2008). The short video (neutral topic manipulated the language mode of the language that followed eliminating L1 influence on the L2 and vice versa on the FLE (Elston-Güttler et al., 2005). For instance, if participants started answering the main questionnaire in English a short clip in their native language was displayed before they moved on to the Greek part of the questionnaire. Comprehension questions aimed to fully engage participants in the task.

The language inducing mode video with the comprehension questions was presented in the beginning of each language survey section. Each language block of the main questionnaire was preceded with a short clip in the target language, to ensure that participants change the internal language model (Grosjean, 2008). Additionally, since this was an online questionnaire with written instructions only in the target language it adds environmental language control to some extent as no oral instructions will be given by a bilingual (Yu & Schiwieter, 2018). Last, the current study collected participants' background information using the LEAP questionnaire in order to investigate what other background factors; the approach and timing of language acquisition, along with cultural connections facilitate or moderate the FLE on moral judgement (Hayakawa et al.,2016).

3.3 General procedure

The four versions of the questionnaire were created in order to counterbalance the presentation of each type of scenario in each language and to make sure that participants were presented with a SP and a NSP dilemma in both languages. Questionnaire versions one and two started in English (FL) and the second part was in Greek (NL) and vice versa for versions three and four.

Participants were assigned to one of the four different versions of the questionnaire consisting of the following parts. Four hypothetical Personal dilemmas [1](require the agent to directly produce harm) deriving from Wong & Ng, 2018, originally brought together in a study by Christensen et al., (2014). Here participants had to state (7-point scale)[2] how likely it would be that they commit the action in the given dilemma, rate how emotionally upsetting they found each scenario and how well they have understood it. Each language block in each version was assigning two scenarios in Greek and two in English.

Then each experiment focused on weighting evidence on the potential role of the proposed mechanisms underlying FLE (Hayakawa et al, 2016) “cognitive overload” and “reduced emotionality” (Kirova & Camacho, 2021). Experiment 1 focused on the cognitive load hypothesis where the effects of foreign language proficiency on the FLE were examined, along with the role of foreign language fluency. Experiment 2, on the other hand focused on the reduced emotionality hypothesis and investigated the role of emotionality on the FLE.

Chapter 4. Experiment 1

The FLE on moral judgment and the role of FL proficiency and fluency.

4.1 Introduction

Based on the main theoretical background and rationale analysed above the FLE posits that cognitive processing is altered when individuals use a language other than their native tongue, often leading to more deliberate and reflective thinking (Keysar et al., 2012). This phenomenon has been attributed to increased cognitive load, as non-native speakers spend more cognitive resources to process information in a foreign language (Costa et al.,2014). However, the role of language proficiency and disfluency in modulating the FLE remains contentious (Hayakawa et al., 2016). Previous studies have largely relied on self-reported proficiency measures, which may not accurately capture the true impact of the cognitive load hypothesis and, consequently, the role of FL proficiency on the FLE. Hence various authors acknowledge the possible limitation and urge for further research using more robust measures (Dylman & Bjärtå, 2019; Marian et al., 2007; Privitera et al., 2023; Stankovic et al., 2022; see also meta-analyses by Circi et al., 2021, and Del Maschio et al., 2022). To address the aforementioned gap, the current experiment of this PhD thesis investigated the cognitive load hypothesis by addressing foreign language proficiency by employing an objective robust and standardised measure.

In addition to the standardized FL proficiency method, fluency will also be measured using semantic and phonemic standard fluency tests, aiming to provide an even clearer picture of each participant's actual fluency level as well. Fluency tasks are a valid, widely used tool that assess vocabulary access ability, offering a basis for comparison between groups of participants (Shao et al., 2014), in this case bilinguals' L1 and L2 fluency. For example, individuals with smaller vocabularies produce fewer words in standard fluency tests (Sauzéon et al., 2011).

Verbal fluency tasks are widely used to assess cognitive and linguistic abilities, particularly in bilingual and multilingual populations, as they measure lexical retrieval efficiency under timed conditions (Bialystok et al., 2008). These tasks are typically divided into category fluency, which requires participants to generate words within a given semantic category (e.g., "animals" or "objects"), and letter fluency, where words must begin with a specific letter, such as "A" or "S" (Kosmidis et al., 2004; Shishkin et al., 2018). Category fluency tasks primarily tap into semantic memory and conceptual organization, whereas letter fluency tasks assess phonemic retrieval and executive functioning, particularly in relation to cognitive flexibility (Henry & Crawford, 2004). Hence, FL fluency can provide more nuanced approach to assess the potential role of cognitive load in the FLE.

Therefore, to assess the potential role of FL proficiency and fluency in moral decision making, experiment 1 first utilizes a comprehensive assessment of participants' language proficiency to ensure a precise evaluation of their language capabilities by employing a standard foreign language proficiency test. Then, it employs standardized fluency tests (category and letter fluency tests) to gauge foreign language fluency, providing an extra comparative analysis to better understand the relationship between cognitive load and the FLE. By implementing these rigorous proficiency assessments, this research seeks to clarify the role of foreign language proficiency on the FLE, thereby contributing to resolving the uncertainty present in previous studies.

Additionally, the current study will investigate the role of the type of dilemmas ‘self-preservation’ and ‘non-self-preservation' on the FLE on Moral Judgment on both experiments. As explained in the theoretical chapter of this PhD these two types of dilemmas might induce different moral decisions and affect the FLE. Therefore, to expand on previous research I will investigate whether dilemmas that involve saving oneself could induce the FLE more in contrast to dilemmas that concern only other individuals (Białek & Fugelsang, 2019) and whether proficiency might modulate this effect

The current study aimed to attain further insights on the role of foreign language proficiency and fluency on the FLE on moral judgement in more depth by investigating:

1. To what degree are moral judgements judged in a more utilitarian manner when using the L2 instead of the L1?
2. What is the role of foreign language proficiency on the FLE on moral judgement?
3. What is the role of foreign language fluency on the FLE on moral judgement?
4. What is the role of the type of dilemma SP/ NSP on the FLE?

Hence, the current study asked participants to morally judge some harmful hypothetical personal moral dilemma dilemmas which require participants to make the utilitarian decision of killing/ sacrificing a person to save more (1-10 people) (Wong & Ng, 2018). Research using moral dilemmas is a valuable 'artificial' methodology that has increased in recent years and offers the fascinating prospect of investigating human moral cognition and its psychological principles (Christensen et al., 2014).

Additionally, the role of the type of dilemma SP/ NSP on the FLE was also explored as the dilemmas were divided into two groups; SP and NSP something that was not acknowledged in previous similar research (Cavar &Tytus,2018; Wong & Ng, 2018). The role of language proficiency (with a more solid proficiency level measure) and emotional response to the dilemmas was also explored (by asking participants to rate the perceived emotionality of each dilemma) in an attempt to weigh evidence towards the emotional reduction hypothesis or the cognitive load hypothesis, or both.

One of the great unknowns in FLE research is whether the effect is rooted in more processing effort resulting from lower proficiency and poorer fluency in the L2. Yet, at the same time FL proficiency has been poorly measured, and fluency has not been measured at all. This leads to the novelty of the current study where instead of using only self-rating proficiency and fluency questionnaires, used in previous studies, this will be the first study to investigate the FLE on moral judgment by providing a concrete representation on participants' current foreign language proficiency level (see 4.2.2.2) and fluency level (see 4.2.2.3) by adding both:

- a standard language proficiency test in the FL (Oxford QPT) providing each participant's formal language skills score complying with the international CEFR scale) (Oxford University Press, n.d.)

- two fluency standard tests (deriving from: Shishkin et al., 2018 and Kosmidis et al. 2004)

What is more, another important aspect is that previous research, mentioned in the meta-analysis by Circi et al., (2021; see also section above) used the dual effect pattern where two groups of different participants were assigned to either the native or the FL condition (known as the dual-effect paradigm) and results from different participants were compared. For this reason, the authors suggest that moral decision-making should be examined on the same participants. The current study has created a within-subjects design, where the FLE will be tested on the same participants in an aim to provide a more accurate representation of the possible FLE by eliminating the uncertainty of results deriving by comparing data between people with different dispositional perspectives and qualities (Hu & Reiterer, 2009). In order to achieve this, a language mode task is used to induce the target language, along with some comprehension questions to fully engage participants in the language mode task.

4.2 Methods

4.2.1 Participants

When designing the first experiment, the mean score of participants was initially calculated based on sample size determination using effect sizes reported in previous, similar research (e.g., Cavar et al., 2018; Geipel et al., 2015; Wong & Ng, 2018). These studies, which included samples such as the 86 participants in Wong & Ng (2018), informed my expectation of a medium effect size (Cohen's $d \approx 0.35$).

To evaluate whether my sample size was sufficient to detect the hypothesized effects, I conducted an analytic power analysis using a paired-samples *t*-test framework. Specifically, I assumed a moderate effect size of Cohen's $d = 0.35$, a significance level (α) of 0.05, and a two-tailed alternative hypothesis. Given that my design involved 107 participants, each providing paired measurements, the data were treated as 107 pairs. The power analysis, performed using the R package *pwr*, yielded a power estimate of 95.2% (power = 0.9519328). This high power indicates that, under these assumptions, my study was more than adequately powered to detect a

mean difference corresponding to an effect size of 0.35. In other words, the probability of a Type II error is low, suggesting that the sample size of 107 participants was sufficient to test the research hypotheses.

Additionally, to further assess whether the sample size was adequate for detecting meaningful effects, a simulation-based sensitivity analysis was conducted using a Monte Carlo approach. This analysis estimated statistical power across a range of effect sizes (Cohen's $f = 0.1$ to 0.5), using 1,000 simulated datasets per effect size and a linear mixed-effects model. Results indicate that the study design provides exceptionally high power—96.5% for small effects (Cohen's $f = 0.1$) and 100% for medium to large effects (Cohen's $f \geq 0.2$). Given that the conventional benchmark for adequate power is 80%, these findings confirm that the current sample size is more than sufficient for detecting small to large effects (see Supplementary Figure S1). The high statistical power is likely attributable to the within-subjects design, which reduces inter-individual variability and enhances sensitivity. These results support the robustness of the study's inferential conclusions, minimizing the risk of Type II errors and ensuring reliable effect detection. Nonetheless, it is worth noting that the current study may be overpowered due to its design, particularly in comparison to previous studies.

Participants in this experiment were required to be adult speakers of Greek as their L1 and English as a FL, with L1 dominance confirmed through self-reported linguistic background data collected via the LEAP-Q and language proficiency assessed through the QPT. To ensure that participants were representative of L1-dominant bilinguals, inclusion criteria required that Greek remained the dominant language in social, familial, and everyday interactions, while English was primarily acquired through formal education rather than in a naturalistic or immersive environment. Participants who reported balanced bilingualism or native-like proficiency in English were excluded, as were individuals who reported early English exposure in a home environment where English was spoken regularly. A specific case involved a participant who reported an age of acquisition (AoA) of 1 year for the FL but was retained in the study because their self-reported language dominance, daily language use, and proficiency measures confirmed that Greek remained their dominant language. This participant's English use was limited to work-related settings, and their QPT score placed them in the low proficiency group, aligning with the study's focus on L1-dominant bilinguals. The final sample consisted of 107 Greek-

English bilinguals taking part in the study (N=107), of which 63 were male and 44 were female, all graduates and staff (at all levels) from local colleges and universities in Cyprus. All participants were adults from a broad range of ages (M_{age}=37, SD_{age}=12) who had acquired English as a foreign language through educational settings (see **Table 2** and **Table 3** for demographic information). Participants voluntarily consented to participate in this study for a chance to win a 40 € voucher. The study was approved prior to data collection by the Faculty of Arts and Social Sciences and Research Management School Research Ethics Committee at Lancaster University, UK (Reference code: **FL20091).**

Table 2

Demographic characteristics of participants

		Count	Column N %
Gender	Male	63	58.88%
	Female	44	41.12%
	Total	107	100.00%
Higher Education Level	Less than High School	0	0.00%
	High School	1	0.93%
	Professional Training	2	1.87%
	Some College	4	3.74%
	College	8	7.48%
	Some Graduate School	12	11.21%
	Masters	64	59.81%
	Ph.D./M.D./J.D.	13	12.15%
	Other	3	2.80%
	Total	107	100.00%

Table 3

Descriptive statistics for continuous demographic variables.

	N	Min	Max	Mean	Std. Deviation
Age	105	20.00	78.00	37.35	12.01
Age of Acquisition in English	96	1.00	18.00	8.07	2.95
Years spent in FL country	81	0.00	40.00	7.13	8.84

4.2.2 Materials.The current study used a within-subjects design where the same participants answered the moral dilemmas in both languages. Four versions of the moral dilemmas' questionnaire were created to counterbalance the presentation of each type of scenario in each language and to make sure that participants were presented with a SP and a NSP dilemma in both languages. Valid responses for each version: Version 1: N=24, Version 2: N= 29, Version 3: N=29, Version 4: N=25). Questionnaire versions one and two started in English (FL) and the second part was in Greek (NL), and vice versa for versions three and four. The anonymous survey was set online via an online survey platform - Lancaster Qualtrics.All items in Greek were translated and back-translated by professional English language teachers whose native language was Greek, to ensure consistency between the Greek and English sections.

4.2.2.1 Language mode inducing task

A language-inducing mode video (neutral topic, ~4 min duration) with comprehension questions was presented at the beginning of each language survey section. Each language block of the main questionnaire was preceded by a short clip in the target language to ensure that participants shifted their internal language mode (Grosjean, 2008). For instance, if they began answering the main questionnaire in English, a short clip in their native language was shown before they moved on to the Greek part of the questionnaire. The comprehension questions were designed to fully engage participants in the task. The main topic of both the English and the Greek videos was on how to manage and organise time more effectively. The theme of the Greek video focused on how to organise the day more effectively, and the theme of the English video focused more on how to prioritise your tasks more effectively. Two comprehension questions were included for each video to assess participants level of engagement and to ensure they

effectively engage with the target language. Participants with empty or wrong responses were excluded from the sample.

4.2.2.2 Moral Dilemma Task

Participants were assigned to one of the four different versions of the questionnaire consisting of 4 moral dilemmas (2 different dilemmas in each language block from which half were SP and the other half NSP). These four hypothetical personal dilemmas (requiring the agent to directly produce harm) derive from Wong & Ng, 2018, originally brought together in a study by Christensen et al. (2014). Following Wong & Ng (2018) methodological steps, participants will read the scenarios followed by a rating scale to indicate the likelihood to commit the action (i.e., to kill one person to save many). The rating scale comprised a 7-point measure with 1 indicating definitively not to commit the action and 7 to definitively commit the action. In addition, participants were asked to rate how upsetting each scenario make them feel (1 = most emotional, 7 = least emotional) and how well they have understood it (1 = not at all, 7 = fully; see Appendix 3b).

The actions included in the hypothetical moral dilemmas were: 1) Pushing an injured person to clear the debris of burning building in order to escape along with five other people, 2). Pushing a large stranger onto the tracks to stop the train and prevent it from killing five workmen, 3) Administer a higher dose of anaesthetic injection that will kill a critically injured patient in order to use the organs to save five other patients, 4) Shoot at an injured diver in order for sharks to stop to eat him and give you and the other nine divers time to escape the sharks and be saved (see Appendix 3a).

Moral dilemmas 1 and 4 are SP (*burning building, shark attack*) as participants will hypothetically save themselves along with others if they choose to commit the action. However, moral dilemmas 2 and 3 (*footbridge, organ transplant*) are NSP dilemmas as they do not involve saving oneself. In all these dilemmas, participants are hypothetically asked to make a utilitarian decision that will result in killing one person in order to save five others.

4.2.2.3 FL proficiency (Oxford QPT)

The selection of an appropriate language proficiency test is crucial in ensuring the validity and reliability of linguistic research. The Oxford QPT was chosen due to its established effectiveness in assessing English proficiency and its strong correlation with the CEFR (Pollitt, 2009). The QPT is a widely recognized tool for assessing English proficiency in second-language learning research. It is also frequently employed by universities, language institutions, and employers to determine the appropriate level of language training for individuals (Trenkic & Hu, 2023).

Therefore, participants' current foreign language English level was measured through the QPT put together by the Oxford University Press and University of Cambridge examinations syndicate representing participants' actual CEFR level of English (Oxford University Press, n.d.) at the time they were taking the test. The test measures knowledge of grammar and the ability to comprehend both explicit and implicit meanings in English words, phrases, and sentences making it particularly suitable for research settings where time constraints must be considered (Khalifa & Weir, 2009). Furthermore, its cost-effectiveness and accessibility ensure that a larger sample of participants can be assessed without the logistical and financial challenges associated with high-stakes proficiency tests (Khalifa & Weir, 2009). Given these advantages, the QPT has been widely utilized in educational and linguistic studies as a practical tool for categorizing proficiency levels in alignment with internationally recognized language standards (Purpura, 2004). By employing this test, the study ensures a standardized, efficient, and valid measurement of participants' English proficiency, which is essential for drawing reliable conclusions.
The QPT is comprised of 60 multiple-choice questions and is time-limited (30 min). Participants are asked to complete as many questions as they can within the timeframe. The test starts with basic structures and vocabulary, and it becomes increasingly difficult by progressing to more complex grammatical structures and vocabulary.

Each participant's QPT score was automatically calculated following the QPT scoring template at the end of each questionnaire through the survey platform used (Qualtrics). The maximal score in the test is 60, but more importantly the range of scores can be translated into the different CEFR levels of proficiency (Oxford University Press, n.d.).

4.2.2.4 Fluency Tasks (Letter & Category in the target language Greek/English)

Four standard category and letter fluency tasks followed the two moral dilemmas in each language block, where participants were asked to name as many words as they could for the broad semantic categories *animals* (Shishkin et al., 2018) and *objects* (Kosmidis et al., 2004) and to produce as many words as possible that started with the letters: A, S, with a time limit of 60 seconds for each fluency task.

The responses were then combined across all four fluency tasks for each participant. Initially, the mean fluency score was calculated across the four tasks to obtain an overall fluency measure for each individual. To further analyse fluency differences, the median fluency score was used as a threshold to split participants into high -and low- fluency groups which will help examine the role of FL fluency on the FLE.

4.2.2.5 LEAP-Q

Last, participants' language background information was collected using the Language Experience and Proficiency Questionnaire (LEAP-Q), which is a validated self-report instrument designed for bilingual and multilingual speakers (age 14 to 80) that collects self-reported information regarding proficiency and language experience data (Kaushanskaya et al., 2020).

This questionnaire (online version available; see Northwestern University Bilingualism and Psycholinguistics Research Group, n.d.) gathers detailed data on participants' language proficiency, exposure, and use across different linguistic contexts (Marian, Blumenfeld, & Kaushanskaya, 2007). Specifically, it measures self-assessed proficiency levels in speaking, reading, and writing for each known language, as well as the age of acquisition, daily exposure, and frequency of use.

The collected data will be analysed using descriptive statistics to summarize participants' language backgrounds and identify potential relationships between language experience and other study variables. Additionally, this information will be used to ensure that participants meet the inclusion and exclusion criteria for the study (e.g., L1 Greek; dominant language).

Furthermore, self-rated proficiency scores from the LEAP-Q will be compared with the Oxford QPT scores to assess the correlation between the two and provide insights into the common reliance on self-reported proficiency in the existing literature body. The urge for more

rigorous proficiency tests is emphasised by various authors (see Stankovic et al.,2022, Del Maschio et al., 2022). An exploratory analysis will also be conducted to determine whether the age of acquisition in English and years spent in an FL country play a role in the FLE.

4.2.3 Experiment procedure

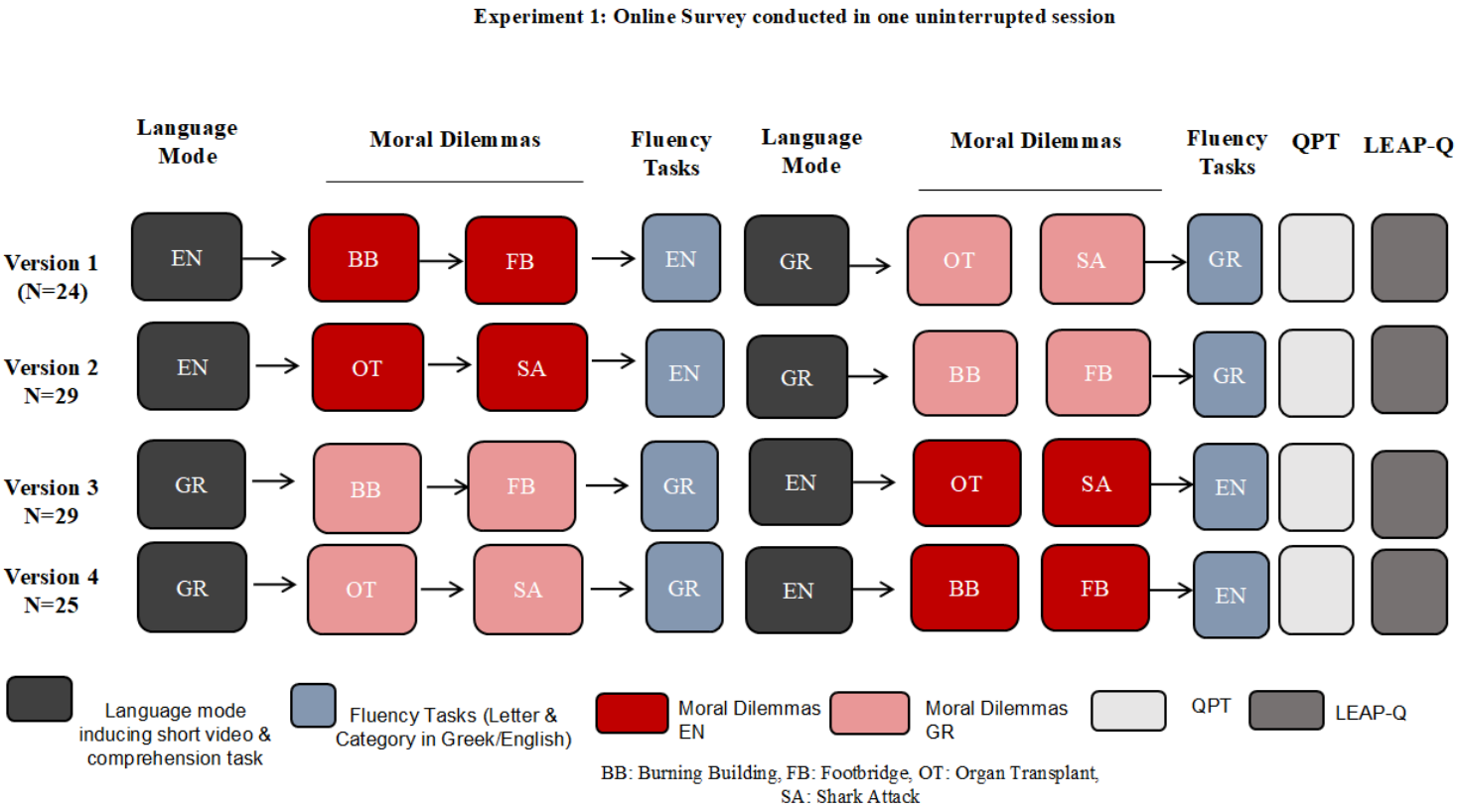
Greek-English participants were assigned to one of the four different versions of the questionnaire illustrated in **Figure 1** consisting of the tasks that were previously explained in detail:

- Language mode task
- Moral dilemmas task
- Fluency tests
- QPT test
- LEAP-Q

The exact sequence of each questionnaire version is demonstrated in **Figure 1**. The online Qualtrics survey was conducted in one uninterrupted session and four different versions were created to counterbalance the four dilemmas across the two languages. As seen in **Figure 1** no participant encountered the same dilemmas in any of the versions. The same participant carried out the same tasks in both languages preceding an inducing language mode. The QPT and LEAP-Q were completed at the end of each version.

Figure 1

Experiment 1: Online Survey *sequence and tasks in each version*



4.3 Results and data analyses: The role of proficiency on the FLE

Participants were divided into two groups based on their proficiency QPT proficiency score. The score was calculated at the end of each survey, and participants were assigned to either the high or low proficiency group. From N=107 participants, N=65 were highly proficient in English (B2, C1, C2 CEFR levels) and N=42 were low proficient participants (A1, A2, B1 CEFR levels).

The distinction between B1 and B2 proficiency levels within the CEFR framework reflects significant differences in linguistic competence and cognitive processing abilities. The B1 level, learners are generally capable of understanding and expressing themselves in familiar

contexts but often struggle with complex language structures and abstract ideas, reflecting limited linguistic autonomy (Hulstijn, 2007). In contrast, B2 learners exhibit greater fluency and flexibility, enabling them to navigate a broader range of topics, including more abstract and specialized discussions, with fewer errors (Khushik & Huhta, 2020). This qualitative leap between the levels not only reflects an expansion in linguistic range but also corresponds to reduced cognitive load during language tasks, as higher proficiency reduces the mental effort required for comprehension and production (Dolgunsöz & Sarıçoban, 2016). According to the official Cambridge exams website, the B1 Preliminary (PET) exam is designed for learners at the B1 level, indicating proficiency in the English language for daily interactions and routines. On the other hand, the B2 First exam (FCE) is targeted at individuals who have achieved a level of English that enables them to function effectively in an English-speaking environment, both in social and professional contexts (Cambridge Assessment English, 2025). Thus, classifying B1 as "low proficiency" and B2 as "high proficiency" aligns with their respective capabilities and the demands placed on cognitive resources during language processing. However, to strengthen the validity of my findings and to ensure that the results were not influenced by mid-proficiency levels, I conducted the same analysis (see section 4.4) using two contrasting groups of proficiency—participants with lower and higher proficiency levels. This approach mitigates the potential confounding effects that mid-level proficiency might have on the research outcomes.

The decision to dichotomize language proficiency was not only based on theoretical and statistical considerations but was also informed by sample size constraints and power analysis. Prior research (e.g., Wong & Ng, 2018; Cavar et al., 2017; Geipel et al., 2015) informed the study's expectations of a small-to-moderate effect size (Cohen's $d \approx 0.35$), leading to a sample size determination of $N = 107$ to ensure sufficient statistical power. Given the study's within-subjects design, a power analysis revealed that this sample size yields a power estimate of 94.8% (power = 0.948), which is well above the conventional 80% threshold. This confirms that the study is adequately powered to detect meaningful within-subject effects.

Since ANOVA requires categorical grouping for between-subject factors, a dichotomous proficiency division could be the most methodologically appropriate choice within this analytical framework (Field, 2018). Furthermore, considering that prior studies have identified threshold

effects in bilingualism (Bialystok, 2017; Kroll & Bice, 2017), dichotomization aligns with previous research that has examined high vs. low proficiency distinctions rather than assuming a strictly linear relationship. Although dichotomization may lead to some loss of granularity, it ensures clear group comparisons and avoids assumptions of linearity, which may not always characterize the relationship between proficiency and cognitive processing (Luk & Pliatsikas, 2016).

Given the study's strong statistical power, I conducted an additional analysis using a continuous approach in R, allowing proficiency to be treated as a gradual predictor. However, since the results were highly similar to those obtained using the mixed-way ANOVA framework, I decided to maintain the dichotomized approach for comparability with previous research—particularly because the study employed a within-subjects design throughout the experiment, rather than the commonly used between-subjects design. While treating proficiency as a continuous variable can offer greater statistical power and preserve meaningful variance (MacCallum et al., 2002), the use of dichotomous groupings in this study aligns with prior research conventions and facilitates interpretation within mixed-way ANOVA designs. Nonetheless, future research may benefit from integrating both categorical and continuous approaches to fully explore the nature of proficiency-related effects (Luk & Pliatsikas, 2016).

4.3.1 FL proficiency and moral decisions

A mixed ANOVA design analysis was used to statistically analyse the data of the current study as presented below. Language (FL: English, L1: Greek) and the type of dilemma (SP=self-preservation, NSP=non-self-preservation) variables were treated as repeated measures that were used in the analysis as the within subject variables. Proficiency level was the independent variable that was used in the analysis as the between-group subject variable (low and high). The dependent variable is the 'commit the action' score (how willing was each participant to commit the action specified in each of the dilemmas).

All effects are reported as significant at at $\alpha = .05$. As predicted, there was a significant main effect of the language on committing the action score by participant, $F(1, 105) = 18.90$, $p <$

.001, η_ρ^2 = .15 (**Figure 2a**), revealing a FLE with 'commit the action' scores significantly higher in English (FL) than in Greek (NL). There was also a significant main effect of the type of dilemma (SP and NSP) on committing the action score, F(1, 105) = 65.50, p < .001, η_ρ^2 = .38, showing that SP dilemmas revealed significantly higher commit the action score than the NSP condition (**Figure 2b**). The interaction between these two factors was not significant, F(1,105) = 3.279, p = .073, ηp^2 = .030 , suggesting that hypothetically saving oneself or not did not yield meaningful different actions across FL an L1.

Additionally, there was a significant interaction effect between the type of dilemma and the proficiency level of the participant, F(1, 105) = 5.05, p =.027, η_ρ^2 = .05. For the SP type of dilemmas, low proficiency level revealed higher committing the action score in contrast to higher proficiency level, t(105)=2.61, p = .010. Whereas for the Non-SP condition, low proficiency level did not reveal a significant difference in committing the action scores compared to higher proficiency level, t(105)=0.38, p = .703, **Figure 3** The other main effects and interactions did not reach significance levels. In order to verify the role of Language proficiency on moral judgement the same statistical analysis was conducted comparing the 25 most proficient and 25 least proficient cases. The results duplicated the main analysis results, where critical interaction between proficiency and language is not significant, ruling out any proficiency effects and establishing that the FLE is present in both high and low proficiency participants.

Figure 2

Mean plots with 95% confidence intervals error bars of committing the action score between language levels (A) and between the type of dilemma (B).

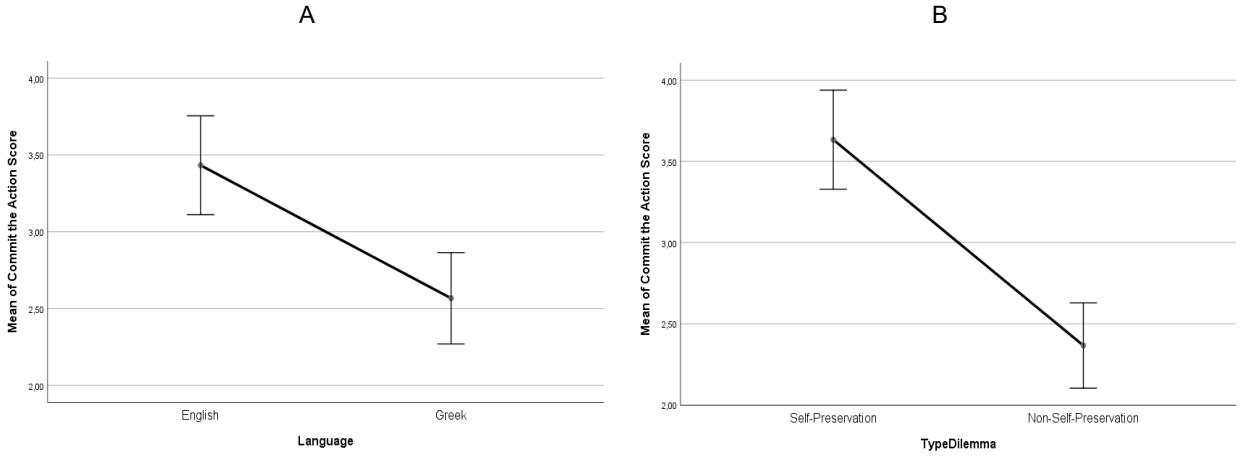
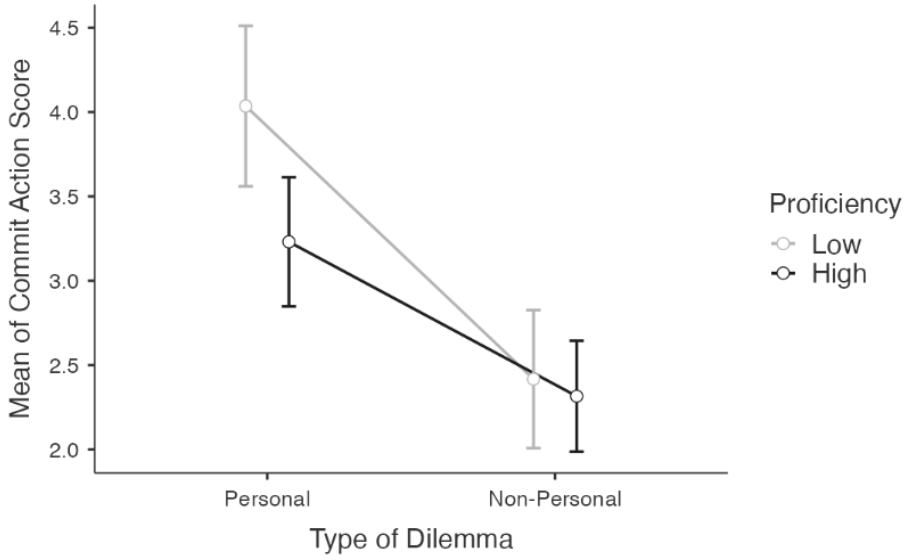


Figure 3

Mean plots with 95% confidence intervals error bars of committing the action score between proficiency levels and type of dilemma.



4.3.2 FL proficiency and language emotionality

After replicating the same method from the previous analysis, but this time with the emotionality score as the dependent variable, the results revealed that there was a non-significant effect of the language on emotionality score by participant, $F(1, 105) = 0.71$, $p = .403$, , $\eta\rho^2 = .01$, suggesting that emotionality score in English (FL) was not significantly different than in Greek (L1). However, there was a significant main effect of the type of dilemma (SP and NSP) on emotionality score, $F(1, 105) = 6.32$, $p = .013$, $\eta\rho^2 = .06$, revealing that SP dilemmas showed significantly more emotional responses than the NSP condition. We also found a marginal interaction between proficiency and type of dilemma on emotionality scores, $F(105) = 3.66$, $p = .058$, $\eta\rho^2 = .03$, but post-hoc comparisons did not reveal any significant effect between low and

high proficiency groups in either SP ($t(105) = 0.63$, $p = .53$), or NSP dilemmas ($t(105) = 1.35$, $p = .177$). All other main effects and interactions were non-significant.

4.3.3 FL proficiency and FL comprehension

The same analyses was conducted with comprehension score as the dependent variable. There was a significant interaction effect between language and the type of dilemma, $F(1, 105) = 4.39$, $p = .038$, $\eta\rho^2 = .04$, see Figure 3. Further comparisons showed that the comprehension score for English and Greek dilemmas did not significantly differ for SP type of dilemmas, $t(106) = 0.54$, $p = .588$, but significantly higher comprehension scores for NSP dilemmas where found for Greek compared to English, $t(106) = 2.77$, $p = .007$. Language × type of dilemma × proficiency level interaction was marginally significant, $F(1,105) = 3.835$, $p = .053$, n.s. $\eta\rho^2 = .04$. Post-hoc comparisons showed that for NSP dilemmas comprehension scores for English were significantly lower than for Greek but only for the low proficiency group ($t(41) = -3.11$, $p = .003$ and $t(64) = -0.98$, $p = .331$, respectively). None of the groups showed significant differences in comprehension across languages in SP dilemmas ($Fs<1$, $ps>.49$). Overall, these results suggest that high proficient bilinguals did not have more difficulty in understanding the dilemmas in their FL as expected, but that the low proficiency group struggled to understand the dilemmas involving saving oneself to a greater degree when this was presented in their FL. Interestingly, this was not the case when the dilemma involved saving others. The other main effects and interactions did not reach significant levels.

Figure 4

Mean plots with 95% confidence intervals error bars of committing the action score between language levels and type of dilemma

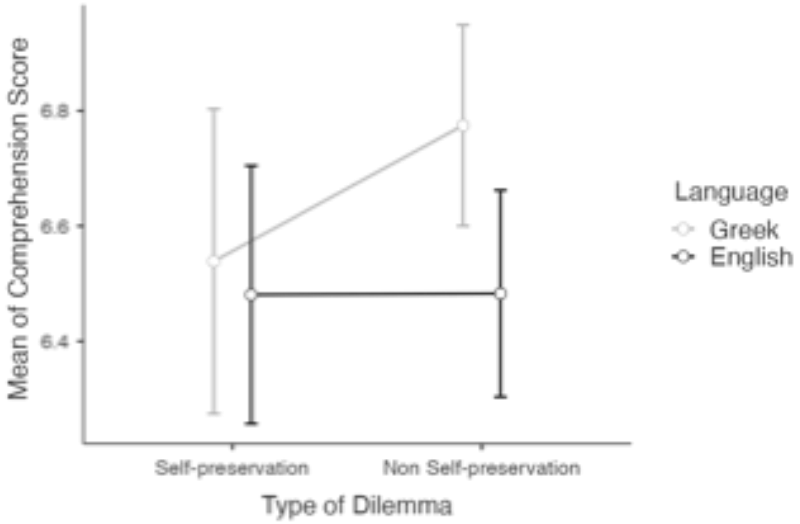
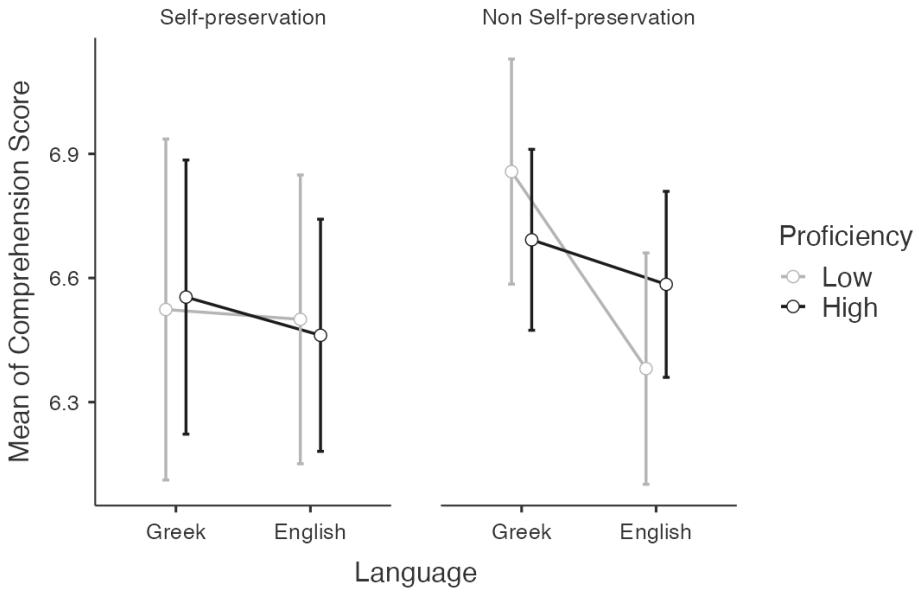


Figure 5

Mean plots with 95% confidence intervals error bars of comprehension score between language levels, type of dilemma and proficiency.

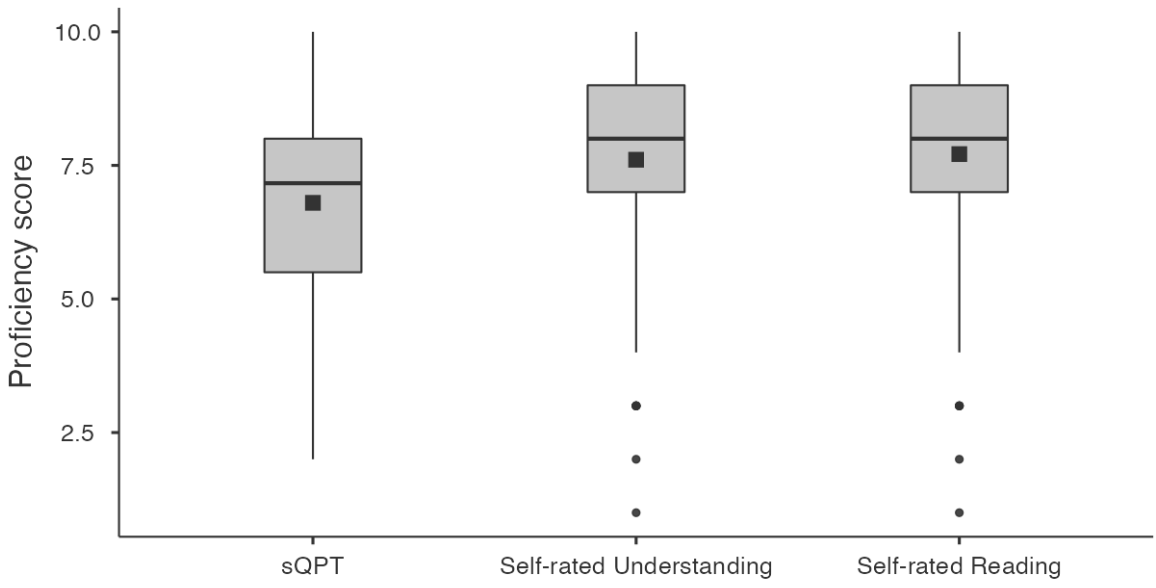


4.4 Proficiency Score vs Self rated proficiency

The proficiency score in the QPT was scaled (divided by six), so that all proficiency measures could be comparable taking values from 1 to 10. Paired-samples t-test were conducted to compare proficiency score in sQPT with self-rated understanding of spoken FL and with self-rated reading proficiency. There was a significant difference in the scores between sQPT (M=6.83, SD=1.89) and the self-rated proficiency in understanding spoken language (M=7.61, SD=1.96; $t(103) = -4.64$, $p < .001$). There was also a significant difference in the scores for sQPT (M=6.83, SD=1.89) and the self-rated reading proficiency (M=7.71, SD=1.90; $t(103) = -5.37$, $p < .001$). In both cases, sQPT has a significantly lower mean value than the mean self-rating for understanding spoken language and reading in the FL. Nonetheless, sQPT was highly correlated with both understanding ($r = .60$, $p < .001$) and reading self-proficiency scores ($r = .61$, $p < .001$, **Figure 6**). This is consistent with the hypothesis that FL users tend to inflate their self-estimates of proficiency in the FL and that formal proficiency measures may provide a more accurate picture of FL proficiency.

Figure 6

Mean boxplots with 95% confidence intervals error bars of proficiency score between self-rated proficiency levels, for reading and understanding scores. Middle lines represent the median and black square the mean.



4.5. Contrasting proficiency groups on the FLE

To check that my results were not driven by mid-proficiency participants and further test the role of FL proficiency on the FLE, I repeated the analysis on two contrasting proficiency groups (of the same participants) by selecting the top 25 participants (23.4%) and bottom 25 participants (23.4%) based on their proficiency scores.

4.5.1 Contrasting cases FL proficiency and moral decisions

A mixture of between group and repeated-measures variables is called a mixed design (Field, 2009). Therefore, a mixed ANOVA design analysis was used to statistically analyse the data of the current study as presented below. Language (FL: English, L1: Greek) and the type of dilemma (SP=self-preservation, NSP=non-self-preservation) variables are treated as repeated measures that will used in the analysis as the within subject variables. Proficiency level is the independent variable that will used in the analysis as the between-group subject variable (low and high). The dependent variable is the 'commit the action' score (how willing was each participant to commit the action specified in each of the dilemmas).

The following coding has been used for this analysis: 1 for English and 2 for Greek (last category is the reference category, so Greek), 1 for self-preservation type of dilemmas and 2 for non-self-preservation type of dilemmas (last category is the reference category, so non-self-preservation), 0 for low proficiency score and 1 for high proficiency scores (post hoc first as control, so low proficiency).

Table 4 demonstrates the descriptive statistics for the Commit the action score within language and type of dilemma condition between proficiency level categories.

Table 4

Descriptive statistics for the commit the action score within language and type of dilemma between proficiency levels.

Proficiency Level	Language	Type of dilemma	N	Mean	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
Low	English	SP	25	4,57	0,29	3,98	5,16
		NSP	25	2,88	0,26	2,36	3,41
	Greek	SP	25	3,50	0,36	2,77	4,23
		NSP	25	1,95	0,24	1,47	2,44
High	English	SP	25	3,69	1,96	3,21	4,18
		NSP	25	2,58	1,94	2,10	3,07
	Greek	SP	25	2,77	1,89	2,30	3,24
		NSP	25	2,05	1,44	1,69	2,40
Total	English	SP	50	4,04	0,19	3,66	4,42
		NSP	50	3,06	0,20	2,65	3,46
	Greek	SP	50	2,70	0,18	2,35	3,05
		NSP	50	2,01	0,14	1,73	2,29

There was a significant main effect of the language on committing the action score by participant, $F(1, 48) = 8{,}151$, $p < .006$, $\eta\rho^2 = .15$ (**Figure 7**) This corroborates the FLE, with more utilitarian responses when making moral judgments in the FL.

There was also a significant main effect of the type of dilemma (SP and NSP) on committing the action score, $F(1, 48) = 33.684$, $p < .000$ (**Figure 8**) , suggesting that participants tend to make more utilitarian choices when the action involved saving themselves. There was no significant main effect of proficiency level, indicating that commit the action scores from low and high proficiency level participants were generally the same, $F(1, 48) = 1{,}28$, $p = .262$, n.s., $\eta\rho^2 = .03$ (**Figure 9**).

Figure 7

Mean plots with 95% confidence intervals error bars of committing the action score between language levels.

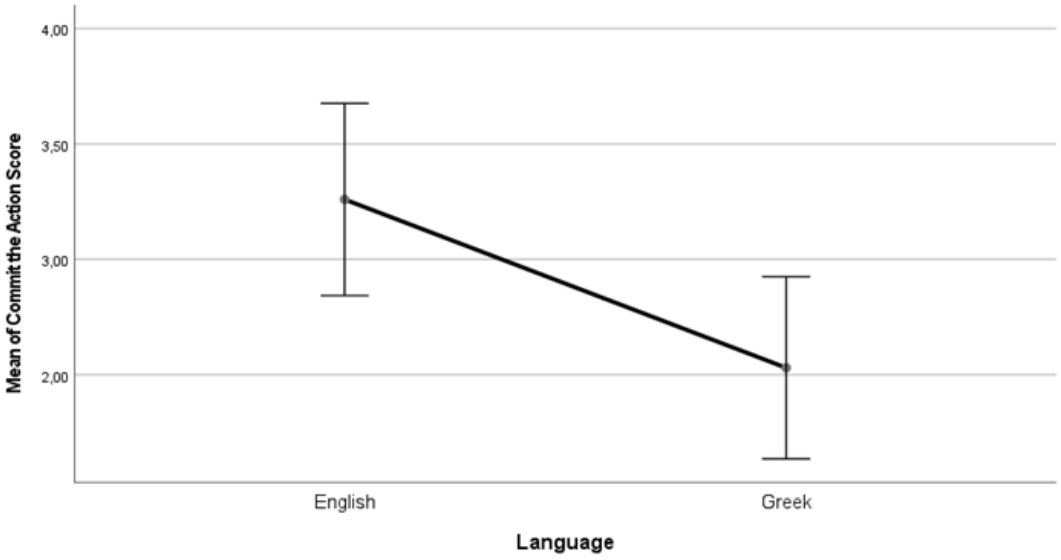


Figure 8

Mean plots with 95% confidence intervals error bars of committing the action score between type of dilemma levels.

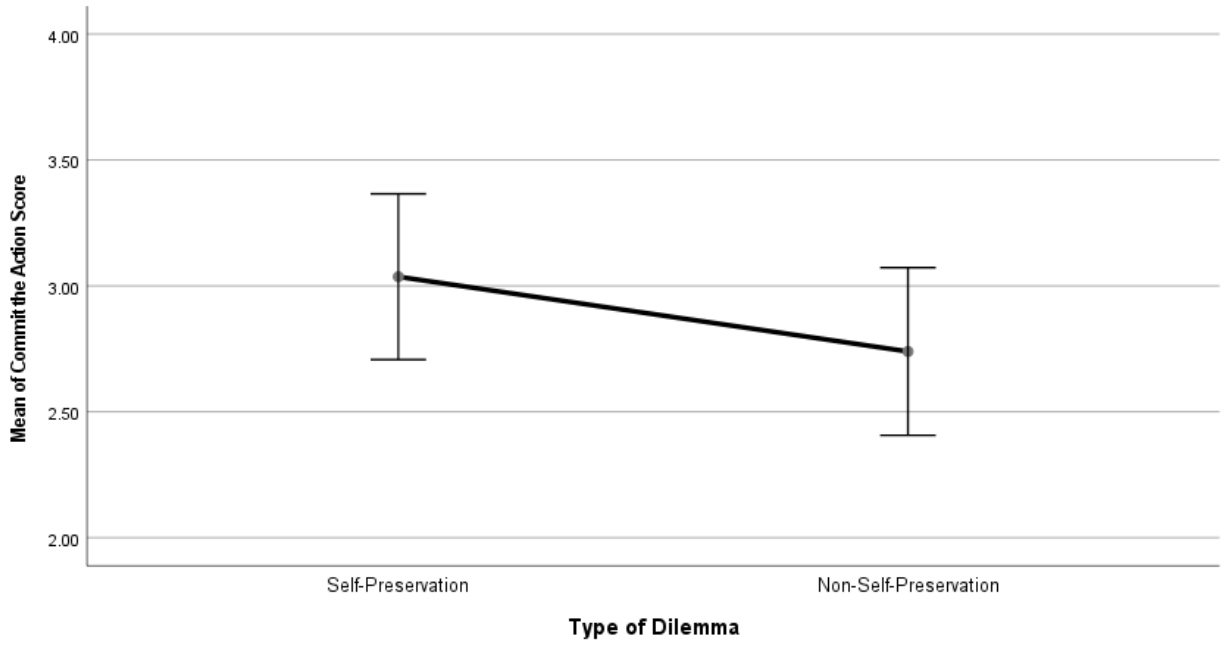
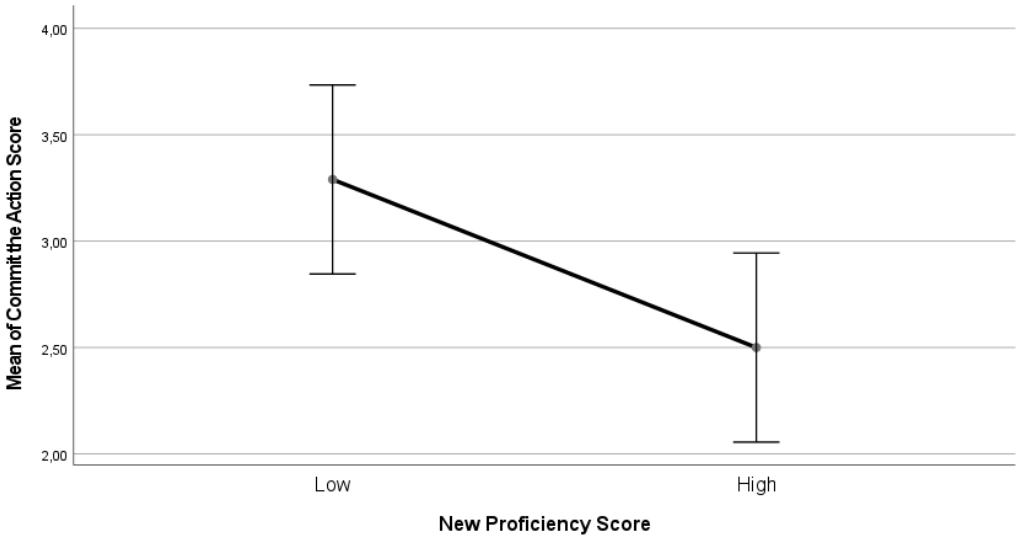


Figure 9

Mean plots with 95% confidence intervals error bars of committing the action score between new proficiency levels.



There was a non-significant interaction effect between the level of language and the proficiency level of the participants, $F(1, 48) = 1{,}28$, $p = .262$, n.s., $\eta\rho^2 = .03$. This indicates that the commit the action score in the two different languages does not statistically differ in low and high proficiency levels. Additionally, there was a non-significant interaction effect between the type of dilemma and the proficiency level of the participant, $F(1, 48) = 0.595$, $p = .444$, n.s. $\eta\rho^2 = .02$. This indicates that the commit the action score did not vary according to their condition.

There was a non-significant interaction effect between the participant's language condition level and the type of dilemma level, $F(1, 48) = 1.380$, $p = .246$, n.s. $\eta\rho^2 = .03$. This indicates that the commit the action score of a different language does not vary in the self-preservation and the non-self-preservation condition. Last, language × type of dilemma × proficiency level interaction was not significant, $F(1,48) = 0.043$, $p = .837$, n.s. $\eta\rho^2 < .01$.

4.4.2 Contrasting cases FL proficiency and language emotionality

Replicating the same method from the previous analysis, a mixed design ANOVA was used, but this time with the emotionality score as the dependent variable. Likewise, Language (English: FL, Greek: L1) and type of dilemma (SP=self-preservation and NSP=non-self-preservation) are the within subject variables, treated as repeated measures. The independent variable proficiency level (low and high) was used as the between-group subject variable in the analysis.

The following coding was used for this analysis: 1 for English and 2 for Greek (last category is the reference category, so Greek), 1 for self-preservation and 2 for non-self-preservation (last category is the reference category, so non-self-preservation), 0 for low and 1 for higher proficiency (post hoc first as control, so low proficiency).

Table 5 presents the descriptive statistics for the emotionality score within language and type of dilemma condition between proficiency level categories.

Table 5

Descriptive statistics for the emotionality score within language and type of dilemma between proficiency levels.

Proficiency Level	Language	Type of dilemma	N	Mean	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
Low	English	SP	25	2.88	1.69	2.18	.34
		NSP	25	2.96	1.74	2.24	.35
	Greek	SP	25	2.84	1.99	2.02	.40
		NSP	25	3.52	2.29	2.57	.46
High	English	SP	25	2.56	1.73	1.84	.35
		NSP	25	2.72	1.57	2.07	.31
	Greek	SP	25	2.44	1.58	1.79	.32
		NSP	25	2.76	1.81	2.01	.36
Total	English	SP	50	2.72	1.70	2.24	.24
		NSP	50	2.84	1.65	2.37	.23
	Greek	SP	50	2.64	1.79	2.13	.25
		NSP	50	3.14	2.08	2.55	.29

There was no significant main effect of the language on emotionality score by participant **(Figure 10)**, F(1, 48)=0.175, p= .678, η_{ρ}^{2} <.00 n.s. suggesting that emotionality score in English (FL) was not significantly different than in Greek. Neither there was a significant main effect of proficiency level (**Figure 11)**, indicating that emotionality score from low and high proficiency level participants were not statistically different, F(1, 48) = 1.180, p= .283, n.s., η_{ρ}^{2} = .02. However, there was a significant main effect of the type of dilemma (self-preservation and non-self-preservation) on emotionality score, F(1, 48) = 4.295, p < .05, η_{ρ}^{2} = .08 (**Figure 12**:where lower emotionality scores represent more emotional responses (rating scale 1-7: 1 most emotional, 7 least emotional), revealing that self-preservation dilemmas were rated significantly more emotional than the non-self-preservation dilemmas.

Additionally, there was a non-significant interaction effect between the level of language and the proficiency level of the participant, $F(1, 48) = 0.325$, $p = 0{,}411$, n.s. $\eta\rho^2 < .01$. This indicates that the emotionality score of different levels of language (L1, FL) does not vary in low and high proficiency levels.

Figure 10

Mean plots with 95% confidence intervals error bars of emotionality score between language levels.

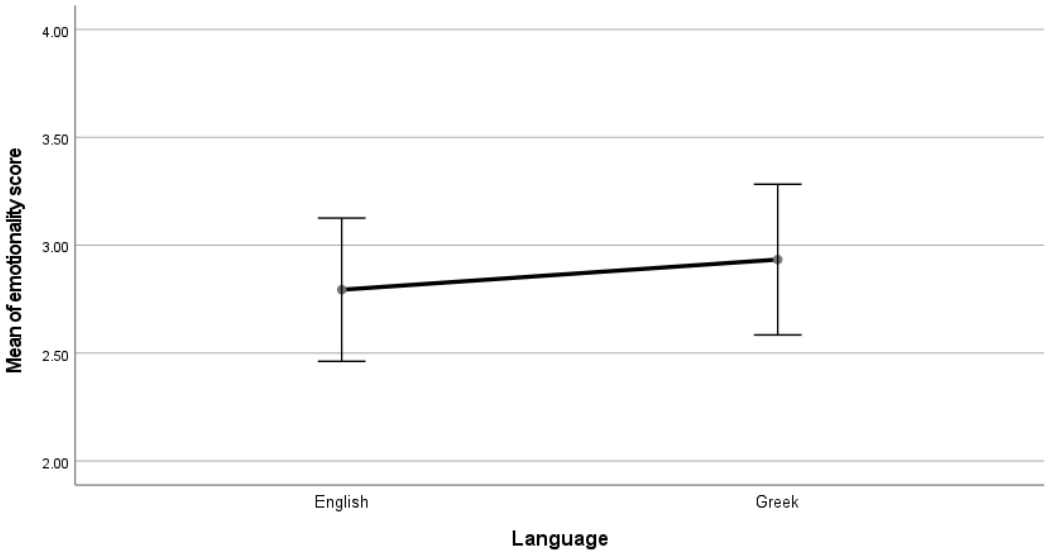


Figure 11

Mean plots with 95% confidence intervals error bars of emotionality score between proficiency levels.

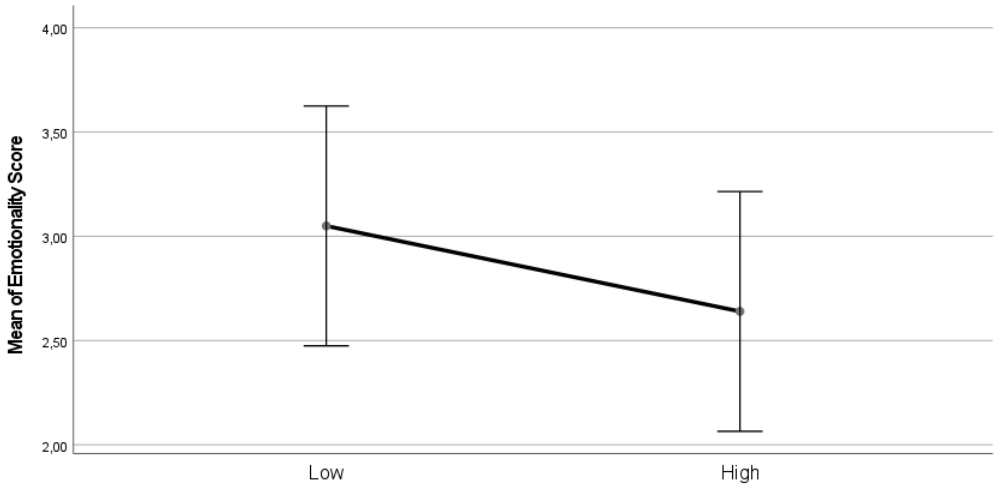
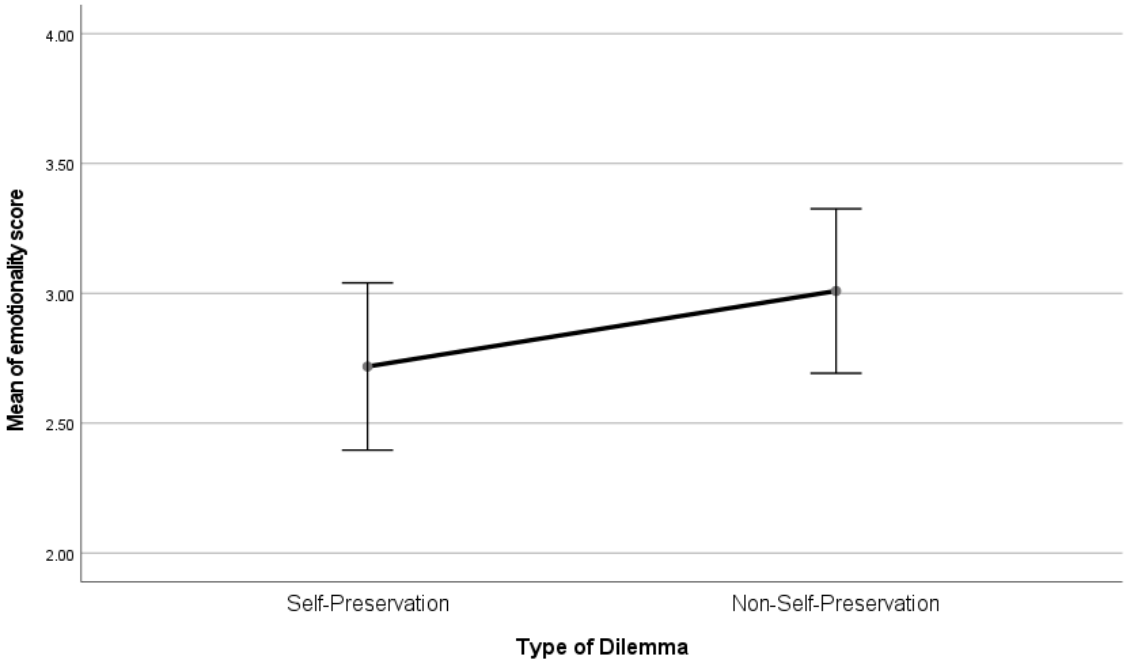


Figure 12

Mean plots with 95% confidence intervals error bars of emotionality score between type of dilemma levels.



There was no significant interaction effect between the type of dilemma and the proficiency level of the participant, $F(1, 48) = 0.219$, $p = .642$, n.s. $\eta_\rho^2 < .01$. This indicates that the emotionality score does not differ according to their condition. Moreover, there was a non-significant interaction effect between the level of language and the type of dilemma level of the participant, $F(1, 48) = 2.320$, $p = .134$, n.s. $\eta_\rho^2 = .05$. This indicates that the emotionality score in a different language does not differ in the self-preservation and the non-self-preservation condition. Last, language × type of dilemma × proficiency level interaction was not significant to explain the emotionality score $F(1, 48) = 0.778$, $p = .382$, n.s. $\eta_\rho^2 = .02$.

4.4.3 Contrasting cases FL proficiency and FL comprehension

A mixed design ANOVA analysis was used and is presented below for the dilemma comprehension score. Language (English and Greek) and the type of dilemma (SP=self-preservation and NSP=non-self-preservation) variables were treated as repeated measures that were used in the analysis as the within subject variables. Proficiency level is the independent

variable that was used in analysis as the between-group subject variable (low and high). The dependent variable is the comprehension score.

The following coding was used for this analysis: 1 for English and 2 for Greek (last category is the reference category, so Greek), 1 for self-preservation and 2 for non-self-preservation (last category is the reference category, so non-self-preservation), 0 for low and 1 for higher proficiency (post hoc first as control, so low proficiency).

Following **Table 6** presents the descriptive statistics for the comprehension score within language and type of dilemma condition between proficiency level categories.

Table 6

Descriptive statistics for the comprehension score within language and type of dilemma between proficiency levels.

Proficiency Level	Language	Type of dilemma	N	Mean	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
Low	English	SP	25	6,50	0,14	6,21	6,79
		NSP	25	6,38	0,14	6,09	6,67
	Greek	SP	25	6,52	0,20	6,12	6,93
		NSP	25	6,86	0,09	6,68	7,03
High	English	SP	25	6,46	0,16	6,15	6,77
		NSP	25	6,58	0,11	6,36	6,81
	Greek	SP	25	6,55	0,17	6,21	6,90
		NSP	25	6,69	0,13	6,43	6,95
Total	English	SP	50	6,48	0,11	6,26	6,69
		NSP	50	6,50	0,09	6,33	6,68
	Greek	SP	50	6,54	0,13	6,29	6,80
		NSP	50	6,76	0,09	6,59	6,93

There was no significant main effect of the language on comprehension score by participant, F (1, 48) = 1,552, p= .219, n.s.. Contrasts revealed that score in English (FL) was not significantly different than in Greek (L1) , F(1, 48) = 1,552, p= .219, n.s., $\eta\rho^2$ =.03. There was a non- significant main effect of the type of dilemma (self-preservation and non-self-preservation) on the comprehension score, F(1, 48) = 1.970, p = *.167,* n.s indicating that participants' comprehension scores did not differ significantly based on the type of dilemma presented.

Contrasts revealed that SPdilemmas have no significantly different comprehension score than the NSP condition, $F(1, 48) = 1.970$, $p = .167$, n.s, $\eta\rho^2 = .04$.

There was no significant main effect of proficiency level, indicating that comprehension score from low and high proficiency level participants were approximately the same, $F(1, 48) = 1{,}226$, $p = .274$, n.s., $\eta\rho^2 = .02$.There was a non-significant interaction effect between the level of language and the proficiency level of the participant, F(1,48) = 1,226, p= .274, n.s., $\eta\rho^2 = .02$. This indicates that the comprehension score of different levels of language was not differed in low and high proficiency levels. Also there was a no significant interaction effect between the type of dilemma and the proficiency level of the participant, $F(1, 48) = 1{,}261$, $p = .267$, n.s. $\eta\rho^2 = .03$. This indicates that comprehension score does not differ according to their condition. Moreover, there was a significant interaction effect between the level of language and the type of dilemma level of the participant, $F(1, 48) = 1.584$, $p = .214$ n.s. $\eta\rho^2 = .03$. This indicates that the comprehension score of different language does not change in the self-preservation and the non-self-preservation condition. However, there was a significant interaction between language × type of dilemma × proficiency level on the comprehension score, F(1,48) = 4.849, p = .032, n.s. $\eta\rho^2 = .09$. Contrasts revealed that in the low proficient group, NSP dilemmas comprehension scores for English were significantly lower than for Greek (t(24) = -2.35, p = .028), but not for SP dilemmas t(24) = -0.43, p = .671, (**Figure 13**). For the high proficient group, neither NSP or SP showed a FLE (t (24) = 0.19, p = .852 and t(24) = -0.57, p = .574, respectively (**Figure 14**).

Figure 13

Mean plots with 95% confidence intervals error bars of comprehension score between type of dilemmas and language for the low proficiency group.

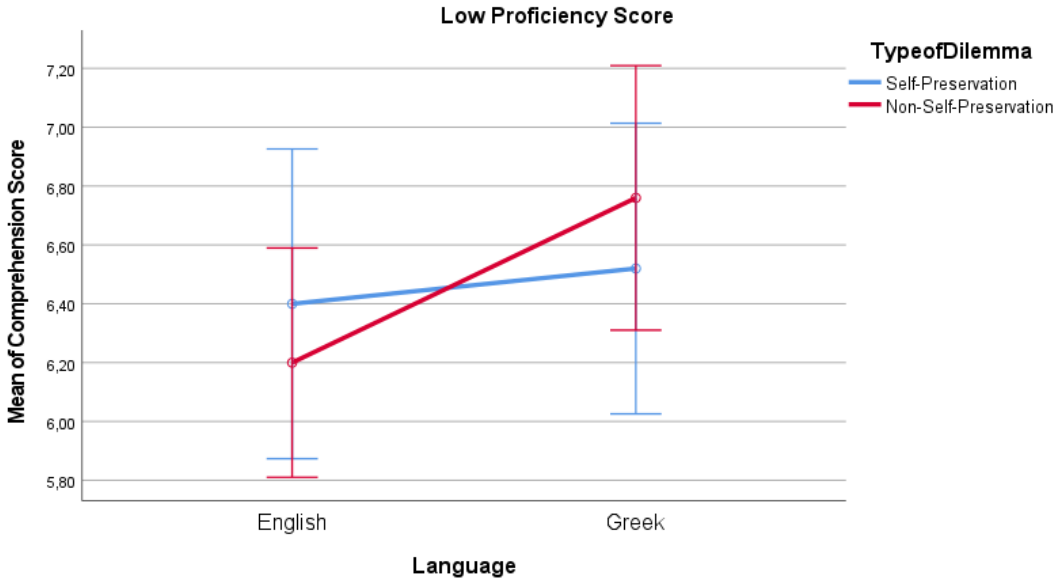


Figure 14

Mean plots with 95% confidence intervals error bars of comprehension score between type of dilemmas and language for the high proficiency group.

To sum up, this results section presented experiment 1 results with a clear FLE on moral judgement, but with no significant influence from FL proficiency on the FLE. The next section will present the results pertaining to the role of foreign language fluency on the FLE.

4.5 The role of FL Fluency on the FLE

Standard fluency tasks were used to measure FL fluency. Participants' fluency scores were divided into two groups the high or low fluency group based on the results from the two letter fluency tasks (where participants produced as many words as possible that start with the letters: A, S) and the two category fluency tasks (where participants named as many words as they could for the broad semantic categories "animals" and "objects") were combined using initially the mean score of the four tasks for each participant. Responses with an average score was less than or equal to the median (8,5 words) were treated as low fluency, whereas those with a score higher than the median were treated as high fluency responses. The four observations that

were equal to the median (8,5) were equally allocated in each group. From N=107 participants N=54 were high fluent participants and N=53 were low fluent participants.

The following coding was used for this analysis: 1 for English and 2 for Greek (last category is the reference category, so Greek), 1 for self-preservation and 2 for non-self-preservation (last category is the reference category, so non-self-preservation), 0 for low and 1 for higher fluency.

Table 7 presents the descriptive statistics for the Commit the action scores within language and type of dilemma condition between fluency level categories.

Table 7

Descriptive statistics for the committing the action score within language and type of dilemma between fluency levels.

Fluency Level	Language	Type of dilemma	N	Mean	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
Low	English	SP	53	3.70	.27	3.15	4.25
		NSP	53	2.58	.24	2.10	3.07
	Greek	SP	53	3.49	.29	2.91	4.07
		NSP	53	2.23	.23	1.77	2.69
High	English	SP	54	4.37	.26	3.85	4.89
		NSP	54	2.81	.26	2.28	3.35
	Greek	SP	54	2.63	.27	2.08	3.18
		NSP	54	1.80	.17	1.46	2.14
Total	English	SP	107	4.04	.19	3.66	4.42
		NSP	107	2.70	.18	2.35	3.05
	Greek	SP	107	3.06	.20	2.65	3.46
		NSP	107	2.01	.14	1.73	2.29

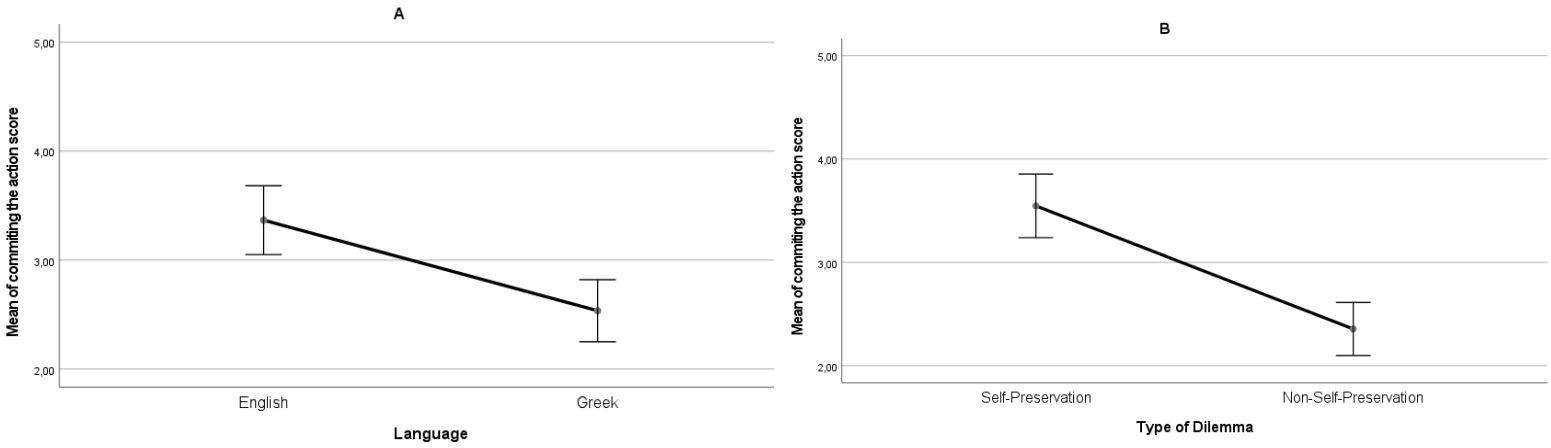
4.5.1 FL Fluency and moral decisions

A mixed design ANOVA analysis was used and presented as follows for the Commit the action score. The language (English and Greek) and the type of dilemma (SP=self-preservation and NSP=non-self-preservation) variables were treated as repeated measures that were used in the analysis as the within subject variables. Fluency level is the independent variable that was used in the analysis as the between group subject variable (low and high). The dependent variable is the commitment score. There was a significant main effect of the language on

committing the action score by participant, F(1, 105) = 19.69, p < .001 (**Figure 15A**), revealing that committing the action scores were significantly higher in English than in Greek. Moreover, there was also a significant main effect of the type of dilemma (SP and NSP) on committing the action score, F(1, 105) = 57.93, p < .001, η_ρ^2 = .36 (Figure 15B), revealing that SP dilemma showed significantly higher committing the action score than the NSP condition.

Figure 15

Mean plots with 95% confidence intervals error bars of committing the action score between language levels (A) and type of dilemma levels (B).



There was a significant interaction effect between the level of language and the fluency level of the participant, F(1, 105) = 8.57, p = .004, η_ρ^2 = .08 that is illustrated graphically in **Figure 16**, indicating that the commit the action score of different levels of language differs in low and high fluency levels. Specifically, the group with high fluency scores showed a significant FLE on commit the action scores. (t(105) = 5.802, p = <.001), but not the low fluency group (t(105) = 1.501, p = .132, n.s.).

Last, language × type of dilemma × fluency level interaction was significant, F (1, 105) = 4.36, p = .039, η_ρ^2 = .04 (**Figure 17**). Planned comparisons revealed that low fluent bilinguals did not show a significant FLE in either SP (t(52) = 0.69 , p = .494) or NSP type of dilemmas (t(52) = 1.47, p = .123), suggesting that language fluency did not significantly influence their moral decision-making in either type of dilemma. However, high fluent bilinguals showed significant higher commit the action scored for English than Greek (FLE) in both SP (t(53) = 4.89, p <.001) and NSP (t(53) = 3.27, p = .002).

Figure 16

Mean plots with 95% confidence intervals error bars of committing the action score between language levels for each fluency level.

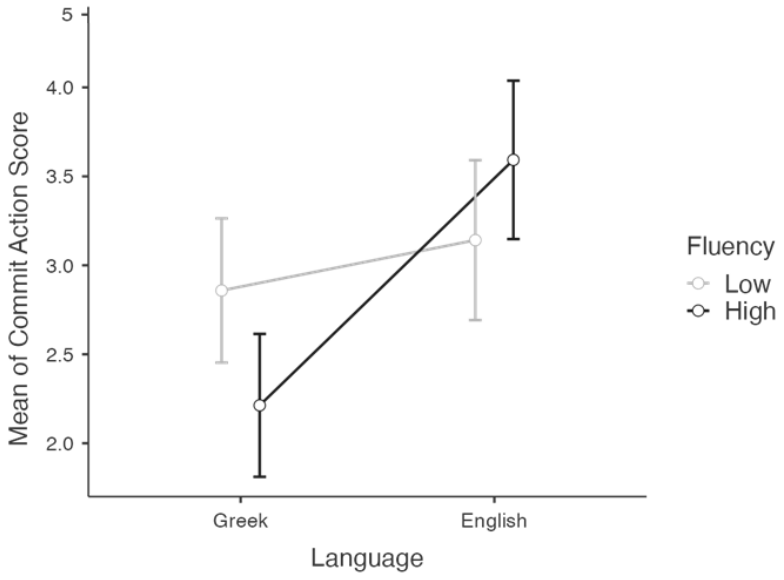
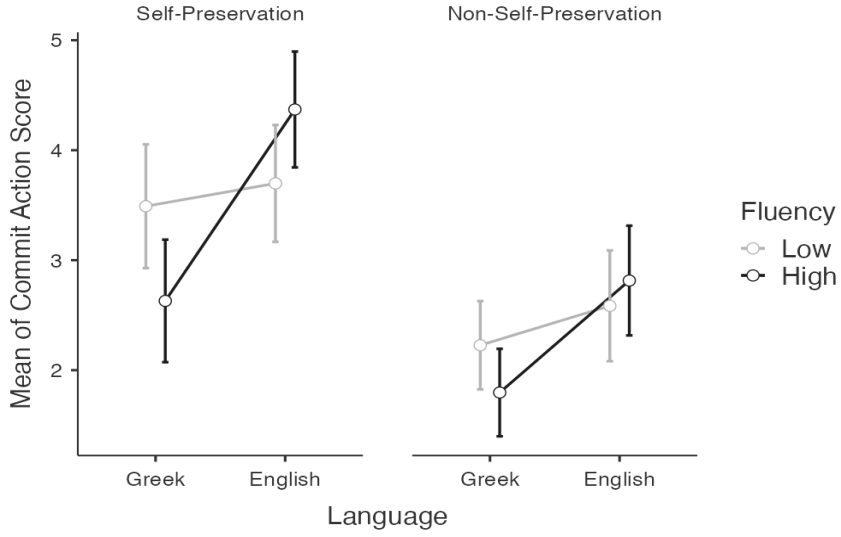


Figure 17

Mean plots with 95% confidence intervals error bars of committing the action score between type of dilemma and language levels, for each fluency level.



4.5.2 Fluency and Emotionality

Once again, a mixed design ANOVA analysis was used and presented for the emotionality score. The language (English and Greek) and the type of dilemma (SP=self-preservation and NSP=non-self-preservation) variables were treated as repeated measures and were used in the analysis as the within subject variables. Fluency level is the independent variable that was used as the between-group subject variable (low and high) in the analysis. The dependent variable is the emotionality score.

There was no significant main effect of the language on emotionality score by participant, $F(1, 105) = 0.38$, $p = .541$, n.s., suggesting that emotionality score in English was not significantly different than in Greek language. However, there was a significant main effect of the type of dilemma (SP and NSP) on emotionality score, $F(1, 105) = 5.53$, $p = .021$, $\eta\rho^2 = .05$, revealing that SP dilemmas showed significantly higher emotionality score which means more emotional responses than the NSP condition. There was also a significant main effect of fluency level, indicating that emotionality score for the low fluency group was significantly higher than the high fluency level participants, $F(1, 105) = 6.86$, $p = .01$, $\eta\rho^2 = .061$. All other main effects and interactions did not reach significant levels

4.5.3 Fluency and FL Comprehension

The analysis revealed no significant main effect of the language on comprehension score by participant, revealing that the comprehension scores in English were not significantly different than in the Greek language. However, there was a significant main effect of the type of dilemma (SP and NSP) on the comprehension score, $F(1, 105) = 4.163$, $p = .044 < .05$ $\eta\rho^2 = .04$, revealing that the SP type of dilemmas have significantly lower comprehension score than the NSP condition. All other main effects and interactions were non-significant.

4.6 Exploratory analysis

Further analysis was carried out to investigate what other variables may underpin the FLE phenomenon (e.g., Athanasopoulos, 2016) that were collected using the participants language background questionnaire (LEAP). To this end, a FLE Index was calculated (responses

from FL were subtracted from those of the NL) and correlation analyses were carried out (see **Table 8**). The exploratory analysis revealed that age, age of acquisition and years spent in an English (FL) speaking country did not drive moral reasoning on the FLE in Greek-English speakers, inconsistent with previous research (Age: Mills & Nicoladis, 2020; age of acquisition: Brouwer, 2019).

Table 8

Pearson Correlations between FLE Index of commit-the-action and emotionality scores with continuous LEAP questionnaire variables

	FLE index Commit the action	FLE Index Emotionality
Age (N = 105)	-0.02	0.10
Age of Acquisition in English (N = 96)	0.13	0.13
Years spent in FL country (N = 81)	-0.08	0.03

Note: Pearson r coefficient values provided, none of the associated p-values were significant. FLE Index is calculated by subtracting English scores from Greek.

4.6.2 Validity of the within-subjects design (experiment 1)

To ensure participants were fully engaged in the intended linguistic mode and to verify the validity of the language induction, a comprehension task was administered immediately following the language-inducing video. Participants who either failed to complete the task or provided incorrect answers were excluded from further analysis. This approach ensured that only those who successfully transitioned into the target language were included in the study, thereby minimizing the risk of cross-language interference. The exclusion of these participants serves as a manipulation check, confirming the effectiveness of the language induction procedure.

Participants received one of four questionnaire variants, each presenting dilemmas arranged differently across languages, so that individuals never encountered the same scenario in both their native and foreign language assessments. This strategy reduced the potential for memory-related bias or repetition effects influencing participants' answers. Additionally, to control for sequence-related biases, the order of language exposure was counterbalanced across

participants: half began with their native language, while the remaining half began with the foreign language. Prior research confirms that counterbalancing efficiently distributes systematic biases across testing conditions, thereby limiting order effects (Brooks, 2012). The integrated approach of varying questionnaire content and employing counterbalanced presentation significantly reduced the risk of carry-over effects.

However, given that no previous studies have tested participants using a within-subjects design on the FLE in a single uninterrupted session, unlike the present PhD, I decided to further assess the effectiveness of this design beyond the methodological safeguards already described. Specifically, I conducted a statistical analysis to examine whether the order in which participants encountered dilemmas (i.e., language block 1 vs. language block 2) influenced moral judgment. To this end, a Mann–Whitney U test was conducted for each language condition separately to determine whether repeating the task influenced participants' moral decisions.

The first analysis compared moral judgment scores for dilemmas encountered in English during the first block ($n = 53$) with those encountered during the second block ($n = 54$). Although the mean for the first-block group ($M = 3.42$, $SD = 1.69$) was slightly higher than that for the second-block group ($M = 3.32$, $SD = 1.64$), this difference was not statistically significant, $t(105) = 0.28$, $p = .778$. This suggests that the block order in which dilemmas appeared in English did not significantly influence moral judgment.

The same analysis was conducted for moral judgment scores in Greek. The comparison involved dilemmas encountered in Greek during the first block ($n = 54$) versus those encountered during the second block ($n = 53$). The mean for the first-block group ($M = 2.67$, $SD = 1.40$) was slightly higher than that for the second-block group ($M = 2.40$), but again, this difference was not statistically significant, $t(105) = 0.92$, $p = .36$. These results suggest that the order in which dilemmas were presented in Greek also did not significantly affect moral judgment.

Importantly, these findings provide further validation for the within-subjects design by demonstrating that participants' moral judgments were not influenced by prior exposure to a similar task in a previous block. Moreover, they indicate that the language-inducing task successfully established the intended linguistic mode, ensuring that responses were not biased by

earlier test exposure. As a result, these findings support the reliability of the study's design in isolating the foreign language effect on moral decision-making.

4.7 Discussion

In these two experiments I predicted that the presentation of moral judgements in a foreign language will lead to a more utilitarian response (e.g., take one life to save five people) than in the L1, where in turn responses are expected to conform to moral imperatives (don't take a life). As predicted, we found a clear FLE with Greek-English bilinguals being more willing to commit the action of killing someone to save five others in their FL than in their L1. It is worth noting that the current study tested the same participants in both of their languages unlike most of previous studies (e.g., Privitera et al., 2023). Nonetheless, my findings are consistent with the majority of previous research that found a FLE in moral judgements (Chan et al. 2016; Cipolletti et al., 2016; Costa et al. 2014; Geipel et al., 2015; Hayakawa et al. 2017; Wong & Ng 2018). Hence, we provide more robust evidence in support of the FLE, where foreign language speakers tend to judge moral decisions less rationally in their L1 and /or less emotionally in the FL (Circi et al., 2021, Geipel et al., 2015, Hayakawa et al., 2016, Costa et al., 2014). Here, we highlight the importance of using within-subjects designs to study the FLE as per to avoid potential confounding variables inherent in between-subjects designs, where different groups of FL speakers are tested in either one language or the other (e.g., Białek & Fugelsang, 2019; Circi et al., 2021; Hu & Reiterer, 2009; Privitera et al., 2023). Controversial or contradictory results in previous literature could indeed be due to comparability between the two samples within the same study.

4.7.1. The role of foreign language proficiency and fluency on the FLE.

The second aim of this study was to investigate the role of foreign language proficiency and fluency on the FLE on moral judgement. Based on previous research (Costa et al., 2014; Geipel et al., 2015), it was predicted that the FLE on moral judgement would be reduced or eliminated in the FL in more proficient/ fluent bilinguals in contrast to less proficient bilinguals. However, as aforementioned the majority of previous research used self-rating proficiency scores

so in order to weigh evidence the role of the cognitive load hypothesis on the FLE the current study rigorously measured foreign language proficiency (Anton et al., 2020). This was done provide a more clear image on the actual role of FL proficiency on the FLE and to fill in the ambiguous literature gap (with results deriving from concrete language proficiency and fluency tests) that exists due to merely self-reported proficiency (Stankovic et al.,2022). Additionally, a recent meta-analysis by Del Maschio et al., (2022) (which is in accordance with Circi et al., 2021) magnitudes the robustness ambiguity of proficiency measures used across 91 experiments. Due to the lack of robust measures, they couldn't either verify or differentiate on the role of proficiency on the FLE stressing out self-report proficiency as a possible factor. In addition, a study by Miozzo et al. (2020) apart from collecting participants' proficiency levels self-rated scores (on a 1–10 scale) they also tested proficiency using an objective grammaticality test. However, participants with low proficiency scores were excluded from the final analysis as only highly proficient bilinguals were considered. Therefore, even though their study used objective proficiency tests for two of their experiments no comparison between low and high proficient participants was made for further validation. In their findings despite of controlling for proficiency, no significant effects were found on moral judgment, indicating that decision-making differences were not influenced by proficiency levels, challenging therefore the cognitive load account- suggesting that the observed shifts toward reduced cognitive biases and increased utilitarian judgments in regional languages were not driven by language mastery, but potentially by sociolinguistic factors, such as the contexts in which these languages are used (Miozzo et al., 2020).

Hence, after measuring language proficiency between high and low proficiency and fluency groups using robust measures the current experiment revealed a reliable FLE regardless of the proficiency of each participant, which goes against the assumption that the FLE is due to the effort that arises from cognitive load (low proficiency) in the FL (Kirova & Camacho, 2021).

To further examine and verify the results of proficiency and eliminate possible biases in the results, deriving from mid-proficiency scores, all the statistical tests were carried out again using two groups of contrasting proficiency cases (25 responses with the lowest proficiency scores and 25 responses with the highest proficiency scores). The results revealed no significant interaction effect between the two languages and the proficiency level of the participants indicating that the FLE was of similar magnitude when contrasting cases of proficiency were

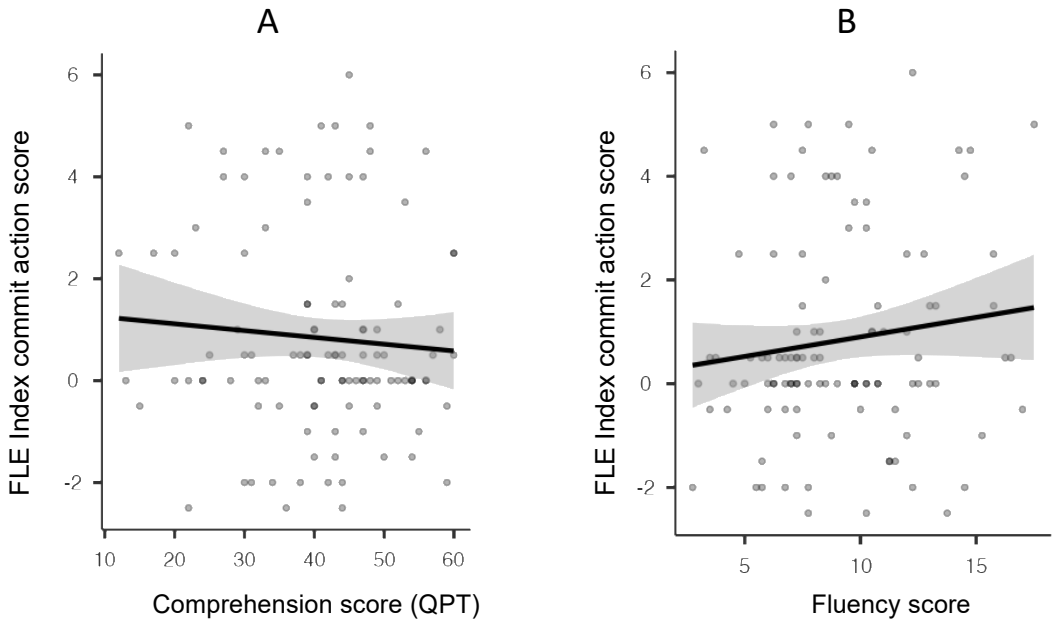
considered, verifying in that way that in the current experiment proficiency (low and high proficiency) is not a factor that mediates or moderates the FLE. This finding could potentially conflict with the results of previous research that claimed that the FLE is due to the cognitive load that occurs from language proficiency barriers (Costa et al., 2014). However, future research could focus exclusively on comparing high and low proficiency groups. This targeted approach would likely offer clearer insights into the specific linguistic features and learning needs characteristic of these distinct proficiency levels.

It is important to note that participants from both low and high proficiency groups indicated approximately the same comprehension score when they had to self-rate how well they have understood the dilemmas. Based on some of the participants' low-level Proficiency scores this could potentially support that in general despite the fact that self-rating data collection is easy it cannot always be reliable in matters of language proficiency and understanding as self-rating participant perceptions could be subjective and inconsistent with the reality (Tomoschuk et al., 2019).

Additionally, fluency level of the participants as measured with a semantic fluency test in English showed a consistent FLE. However, this was only significant for the high fluency group. These results challenged my predictions where higher differences between the two languages were expected for less fluent bilinguals. Nonetheless, as can be observed in Figure 6, these differences could also be driven by the high fluent group being less likely to commit the action in Greek. This thus suggests that other factors inherent in between-group comparisons could potentially be influencing our results. Indeed, when observing fluency and comprehension scores as a continuum in relation to the FLE, no clear linear trend was observed (see **Figure 18A** and **18B**).

Figure 18

Relationship between FLE commit the action scores (English minus Greek) and individual (A) proficiency scores (QPT) and (B) fluency scores. Black lines represent the linear trend and grey areas depict 95% confidence intervals.



When participants were asked to rate how emotional each moral dilemma in each language made them feel no significant differences were recorded in English (FL) and Greek (L1). This is inconsistent with the emotion-reducing hypothesis where foreign language messages usually elicit less significant emotional responses in contrast to the L1 (Dewaele, 2004; Harris, 2004; Harris et al., 2003; Hsu et al., 2015; Iacozza et al., 2017; Zheng et al., 2020). At this point it is important to mention that the current study differs in the design used in contrast to previous research. A within-subjects design (also found in Mills & Nicoladis's, 2020) was used to provide a representation with more increased power by focusing on the variances of the conditions and the experimental control rather than external factors that could arise by comparing different groups of people in a between-subjects design (Białek & Fugelsang, 2019). Therefore, this could potentially raise the question on whether the between-subjects design was appropriate for such experiments and on measuring emotionality. Since I included a language mode induction procedure (with comprehension questions) to reinforce the intended linguistic context, in both experiments, this should have minimized language mixing effects. Additionally, I tested for block effects and found no significant differences, suggesting that order effects did not

systematically impact the results. This indicates that carry-over effects were unlikely to have played a major role in this experiment. However, it is still possible that subtle emotional spillover occurred, which could be further examined in future research. Therefore, in the next chapter the study will explore emotions prior and after the moral dilemma task in order to considering how language effects unfold over the course of an experiment rather than focusing solely on single-item emotionality ratings.

4.7.2 The role of the type of dilemma SP/ NSP on the FLE

Previous authors suggested that people are more willing to commit the action of killing someone when it involves saving themselves (Bloomfield, 2007; Mills & Nicoladis, 2020). The third aim of the study was to investigate the role of the type of dilemma SP/ NSP on the FLE that was merely neglected to be acknowledged in previous research that mainly focused on personal and impersonal dilemmas (Wong & Ng, 2018). The role of the type of dilemma, SP and NSP, was further examined using additional sets of moral dilemmas, along with even more rigorous proficiency measures than those used by Mills & Nicoladis's (2020) study. It is important to mention that as in Mills & Nicoladis (2020) we also used a within-participant design for the type of dilemma, which provides with a more valid representation of responses excluding external barriers that could arise in a between-subjects design that compares different groups of people with potentially different morals (Białek & Fugelsang, 2019). Our results were in accordance with Mills & Nicoladis (2020) study where participants were more willing to commit the action of killing in SP dilemmas (when the agent was at risk too) in contrast to NSP dilemmas (Sachdeva et al., 2015) regardless of the language used. This shows that the type of dilemma is indeed important in moral judgement (Chan et al., 2016).

Less utilitarian responses in NSP dilemmas stimulated less emotionality which explains their deontological decisions (Hayakawa et al., 2016). Yet, SP dilemmas revealed significantly more emotional responses than the NSP condition. Current research has shown that when people are faced with a dilemma that involves saving themselves SP they tend to have more negative emotions, consistent with what we found. A possible explanation for this could be that participants are more sensitive to negative emotions when faced with dilemmas that directly

impact their own survival (Privitera et al.,2023). This could derive from the framework that suggests that utilitarian decisions could be linked to egoistic; actions driven by self-interest and personal benefit (APA Dictionary of Psychology, n.d.); outcome-based decisions (Reynolds & Ceranic, 2007). Another possibility for the more emotional responses could be that guilt is involved in matters of moral norms or hurt done to other people (Hoffman, 2000). What is crucial to note though is that this wasn't modulated by language, which means that the type of dilemmas (more emotional responses in SP) does not seem to be affected by the FLE. As a result, the emotion reducing account (see Keysar et al., 2012) is not consistent with the type of dilemma. If emotion was involved, it would have been expected to find a FLE in SP dilemmas more than in the NSP dilemmas, but we don't. Hence, the role of separated emotion types and emotions in general should be further explored on the FLE regardless of the type of dilemma

Chapter 5. Experiment 2

The FLE on moral judgment and the role of emotions.

5.1 Introduction

The interplay between foreign language emotionality and its impact on the FLE represents a crucial area of exploration. Previous studies have suggested that cognitive load significantly influences the FLE, positing that the additional mental effort required to process a foreign language diminishes emotional intensity and decision-making biases (Hadjichristidis et al., 2015, Costa et al.,2014). However, the findings from my prior experiment contradict the cognitive load hypothesis, as the results demonstrated no significant impact of foreign language proficiency or fluency on the FLE. This unexpected outcome necessitates a re-evaluation of the mechanisms underlying FLE, with a particular focus on the second possible factor that drives the FLE according to research (Hayakawa et al., 2016) the emotional reduction hypothesis.

Recent studies highlight the importance of emotions in driving FLE, indicating that emotional arousal and affective states significantly affect how individuals process a non-native language (Costa et al., 2014; Pavlenko, 2012). Emotions are closely tied to the native language through lifelong experiences and cultural contexts, often leading to reduced emotional resonance when expressed in a foreign language (Dewaele, 2010). This phenomenon, known as emotional distance, can result in more utilitarian and less biased decision-making when using a foreign language due to the reduced emotional intensity (Hayakawa et al., 2017). In light of this evidence, experiment 2 aims to explore the emotional mechanisms underlying FLE more deeply. By examining the different emotional responses elicited by native and foreign languages, this experiment seeks to determine the extent to which emotions, rather than cognitive load, drive FLE.

The design of this experiment is based on strong evidence that emotions significantly influence language processing and decision-making (e.g., Brouwer, 2021; Caldwell-Harris, 2015; Hayakawa et al., 2017; Kyriakou et al., 2023). The emotion-reduced hypothesis suggests that using a foreign language lessens the emotional impact

during language processing and decision-making. This hypothesis argues that emotional experiences and expressions are less intense when conveyed in a non-native language, resulting in more impartial and less biased cognitive processing. The reduced emotional intensity is attributed to weaker affective associations in a foreign language, which is often learned later in life and used less frequently in emotionally charged situations (Dewaele, 2010; Pavlenko, 2012). As a result, individuals tend to make more rational and utilitarian decisions when using a foreign language compared to their native language, since the emotional factors that typically influence decision-making are less pronounced (Costa et al., 2014; Hayakawa et al., 2017). This hypothesis helps explain the FLE, where systematic differences in behaviour and thought patterns emerge when individuals operate in a foreign language rather than their native tongue.

By advancing our understanding of the emotional dynamics involved in the FLE, this chapter seeks to offer a more nuanced explanation of the emotional reduction hypothesis on the FLE by employing robust emotional measures explained in detail in the next section. This understanding can inform practical applications in multilingual contexts, such as international negotiations, bilingual education, and cross-cultural communication, where emotional and cognitive factors intersect in complex ways (Geipel et al., 2015).

The aim of this second study is to attain further insights on the role of foreign language emotionality on the FLE on moral judgement in more depth; Once again in an aim to investigate the literature gap on what drives the FLE proposed by Hayakawa's et. al. (2016) the role of the FL emotional reduction will be examined in order to reveal what is the impact of foreign language emotionality on the FLE on moral judgement. It is expected that moral dilemmas, as they involve physical harm, will instinctively trigger emotional responses that outweigh utilitarian considerations, leading to more controlled, utilitarian responses in the FL (Koenigs et al., 2007). This was investigated by assessing participants' emotional states before and after moral dilemmas in each language using the PANAS-X emotion rating scale (see Horne & Powell, 2016).

5.2 Methods

5.2.2 Participants

To evaluate whether the sample size was sufficient to detect the hypothesized effects, I conducted an analytic power analysis using a paired-samples t-test framework. Specifically, I assumed a small-to-moderate effect size of Cohen's d = 0.35, a significance level (α) of 0.05, and a two-tailed alternative hypothesis. Given that my design involved 106 participants, each providing paired measurements, the analysis treated the data as 106 matched pairs. The power analysis, performed using the R package pwr, yielded an estimated power of 94.6% (power = 0.946). This high level of power indicates that, under these assumptions, my study was more than adequately powered to detect a mean difference corresponding to an effect size of 0.35. In other words, the probability of committing a Type II error is very low, suggesting that my sample size of 106 participants is sufficient for testing my research hypotheses.

Additionally, to further assess whether the sample size (N = 106) was adequate for detecting meaningful effects in my within-subjects design, I conducted a simulation-based sensitivity analysis within the same paired-samples t-test framework. The simulation modelled repeated measures with a within-subject correlation of r = 0.5, reflecting the expected dependency between pre- and post-measurements. I systematically varied effect sizes (Cohen's d) from 0.1 to 1.0 and estimated statistical power across 10,000 replications per condition. The results showed that for a small-to-moderate effect size (d = 0.35), the study achieved 94.6% power, closely aligning with the estimate obtained from the analytic approach using the pwr.t.test () function in R. Moreover, the simulation indicated that even for a slightly smaller effect size (d = 0.3), statistical power remained high (~86%), ensuring sufficient sensitivity to detect moderate within-subject differences (see Supplementary Figure S2). Only for very small effects ($d \leq 0.2$) did power fall below the conventional 80% threshold (e.g., ~53% power for d = 0.2), suggesting that detecting minimal changes would require a larger sample.

Overall, these findings confirm that my sample size of 106 participants provides robust power for detecting meaningful within-subject effects, supporting the validity of

my study design.

One hundred and six (NL; Greek) (FL; English) participants (N=106), of which 49 male and 57 female took part in the study. The participants were mostly graduates, and staff (at all levels), that did not participate in the previous experiment, as demonstrated in **Table 9**, from local colleges and universities in Cyprus. Their mean age is 34 and the mean age they started acquiring their FL (English) was around 8 (**Table 10**).

Table 9

Demographic characteristics of participants

		Count	Column N %
Gender	Male	49	46.2%
	Female	57	53.8%
	Total	106	100.00%
Higher Education Level	Less than High School	1	0.90%
	High School	6	5.70%
	Professional Training	1	0.90%
	Some College	0	0.00%
	College	6	5.7%
	Some Graduate School	11	10.4%
	Masters	66	62.30%
	Ph.D./M.D./J.D.	12	11.3%
	Other	3	2.80%
	Total	106	100.00%

Table 10

Descriptive statistics for continuous demographic variables.

	N	Min	Max	Mean	Std. Deviation
Age	106	20.00	74.00	33.80	9.00
Age of Acquisition in English	106	1.00	16.00	7.71	3.01
Years spent in FL country	34	0.00	40.00	6.00	8.09

5.2.3 Materials

This study employed a within-subjects design, replicating the previous experiment, in terms of design, while testing a new group of participants who had not taken part in experiment 1. Participants completed the questionnaire in both languages to minimize variability caused by individual differences.

5.2.3.1 Language mode inducing task

At the start of each section of the language survey, participants were shown a video designed to activate a specific language mode, followed by comprehension questions. To facilitate a shift in language processing, each segment of the main questionnaire was introduced with a brief video clip in the corresponding language (Grosjean, 2008). For example, if respondents had been answering in English, a short clip in their native language was played before transitioning to the Greek part of the questionnaire. The comprehension questions served to maintain participant engagement and reinforce the language shift.

5.2.3.2 Moral dilemma task

In line with the earlier experiment, all materials in Greek were translated and then retranslated by professional teachers fluent in both Greek and English. This process was undertaken to ensure that the Greek and English versions were consistent and accurately matched. The same four hypothetical personal moral dilemmas were used (see Appendix 3a).

1. *Burning building*: Urging an injured person to clear debris from a burning building to facilitate escape for themselves and five others.
2. *Footbridge dilemma:* Pushing a heavy stranger onto the tracks to halt a train, saving five workers from being killed.
3. *Organ transplant:* Administering a lethal dose of anesthetic to a seriously injured patient to harvest organs and save five other patients.

4. *Shark attack:* Shooting an injured diver to distract sharks, allowing you and nine other divers to escape safely.

The dilemmas were once again counterbalanced into four versions see **Figure 19** *(BB: burning building, SA: shark attack OT: organ transplant, FB: footbridge)* Version 1: English OT, Greek BB version 2: English: BB, Greek: OT, version 3: Greek SA, English FB, version 4**:** Greek: FB**,** English SA

5.2.3.3 Emotional measure: PANAS-X

The PANAS-X (Positive and Negative Affect Schedule) is a scale that aims to measure someone's feelings at the moment and their positive and negative affect. This scale was developed in 1988 and offers a reliable, accurate, and effective method for measuring positive and negative emotions. (Watson et.al,1988).

It consists of 20 affective descriptors, with 10 measuring positive affect (e.g., enthusiastic, active) and 10 evaluating negative affect (e.g., distressed, irritable). Participants rate each emotion on a five-point Likert scale ranging from 'very slightly or not at all' (1) to 'extremely' (5) based on their emotional experience within a specified time frame, such as 'right now'. Positive affect and negative affect scores indicate the intensity of positive and negative emotional states, respectively. High positive affect scores typically reflect greater levels of enthusiasm, energy, and engagement, while elevated negative affect scores suggest higher levels of distress and unpleasant emotions. The PANAS-X is extensively used in psychological research and clinical contexts to assess mood fluctuations, emotional well-being, and responses to stimuli, making it a reliable tool for measuring both state and trait affect (Crawford & Henry, 2004). Due to its adaptability, the scale has been widely applied across various populations and research domains, reinforcing its value in affective science that's why it has been chosen as the emotional measure of the current experiment. Horne & Powell (2016) developed the pre-test and post-test versions of the PANAS-X (including all descriptors in both tests) to assess affective changes in emotions across conditions.

In this experiment, I used the pre-test and post-test versions of the PANAS-X

(originally developed by Watson & Clark, 1994, PANAS-X: Manual for the Positive and Negative Affect Schedule), as adapted into pre-to-post version by Horne & Powell (2016), to measure language emotionality before and after participants were exposed to the moral dilemmas. Participants in this experiment completed the positive and negative affect scales, as well as the guilt, hostility, and joviality scales. Table 11 (found in Horne, & Powell, 2016) divides the exact words that participants rated for each emotion. Hence, each participant will rate their emotional state twice in English (before and after the English moral dilemmas), and twice in Greek (before and after the moral dilemmas in Greek). The presentation order of emotions varied between the pre-test and the post-test.

Having the pre-test baseline for each language allows us to observe emotional states changes for each language independently, whilst at the same time allows us to consider any potential effect of language of operation on PANAS-X ratings during pre-test.

Table 11

Emotion PANAS-X Categories (used in Horne, & Powell, 2016)

Negative affect (10)	afraid, scared, nervous, jittery, irritable, hostile, guilty, ashamed, upset, distressed
Positive Affect (10)	active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud, strong
Hostility (6)	angry, hostile, irritable, scornful, disgusted, loathing
Guilt (6)	guilty, ashamed, blameworthy, angry at self, disgusted with self, dissatisfied with self
Joviality (8)	happy, joyful, delighted, cheerful, excited, enthusiastic, lively, energetic
Note. The number of terms comprising each scale is shown in parentheses.	

5.2.3.4 LEAP-Q

To ensure that participants meet the inclusion and exclusion criteria (such as having Greek as their L1) of the second experiment and to collect language background information for the demographic and descriptive statistics the LEAP-Q was administered (same as experiment 1; Northwestern University Bilingualism and Psycholinguistics Research Group, n.d.) which gathers detailed data on participants'

language use across different linguistic contexts (Marian, Blumenfeld, & Kaushanskaya, 2007).

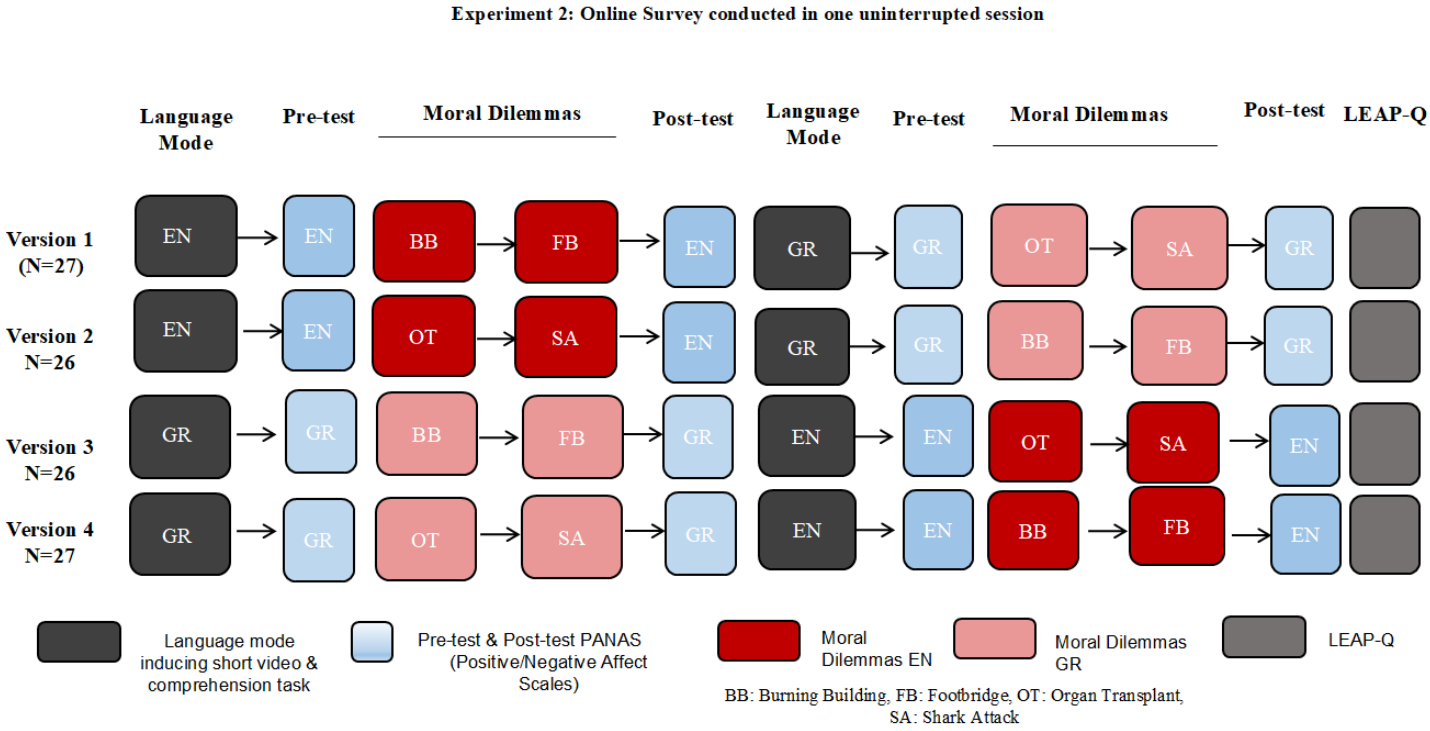
5.3 Experiment procedure

Four versions of the questionnaire were created (see **Figure 19**) in order to counterbalance the presentation of each scenario in each language and to make sure that participants were presented with a SP and a NSP dilemma in both languages. Questionnaire versions one and two will start in English (FL) and the second part will be in Greek (NL) and vice versa for versions three and four. Identically to experiment 1 in the previous chapter and in order to implement a within-subjects design, each section of the language survey began with a video designed to induce the appropriate language mode, followed by comprehension questions. To ensure participants switched their internal language mode (Grosjean, 2008), each language block of the main questionnaire was introduced with a brief clip in the target language. For instance, after answering the main questionnaire in English, participants watched a short clip in their native language before proceeding to the Greek section. The comprehension questions aimed to fully engage participants with the task.

Greek-English participants were assigned to one of the four different versions of the questionnaire below consisting of the following parts (conducted in one uninterrupted session-4 different versions were created to counterbalance the 4 dilemmas across the 2 languages):

Figure 19

Experiment 2: Online Survey Sequence and tasks in each version.



1.1 Language Mode inducing task.

1.2 Pre-test emotion ratings PANAS-X (Appendix 4a): Participants will have to provide a rating of specific positive and negative emotions they are exposed to in the relevant language.

1.3 Participants will be presented with two of the same four hypothetical dilemmas as in the first experiment (One in their L1: Greek and one in their FL: English). Here participants will once again have to state (7-point scale) how likely it would be that they commit the action in the given dilemma. Each language block in each version will be assigning two dilemmas in Greek and two in English and the four dilemmas will be counterbalanced within the four versions.

1.4 Post-test PANAS-X (Appendix 4b): After participants had been exposed to the Moral Dilemmas and hypothetically indicated if they would commit the action,

they will have to do the Post-test PANAS-X where they will have to provide a rating of specific positive and negative emotions in the respective language block.

1.5 Language Mode inducing task.

1.6 Pre-test emotion ratings PANAS-X (Appendix 4a): Participants will have to provide a rating of specific positive and negative emotions they are exposed to.

1.7 Participants will be presented with one of the same four hypothetical dilemmas as in the first experiment (One in their L1: Greek and one in their FL: English). Here participants will once again have to state (7-point scale) how likely it would be that they commit the action in the given dilemma. Each language block in each version will be assigning one dilemma in Greek and one in English and the four dilemmas will be counterbalanced within the four versions.

1.8 Post-test PANAS-X (Appendix 4b): After participants had been exposed to the Moral Dilemma and hypothetically indicated if they would commit the action, they did the Post-test PANAS-X where they provided a rating of specific positive and negative emotions.

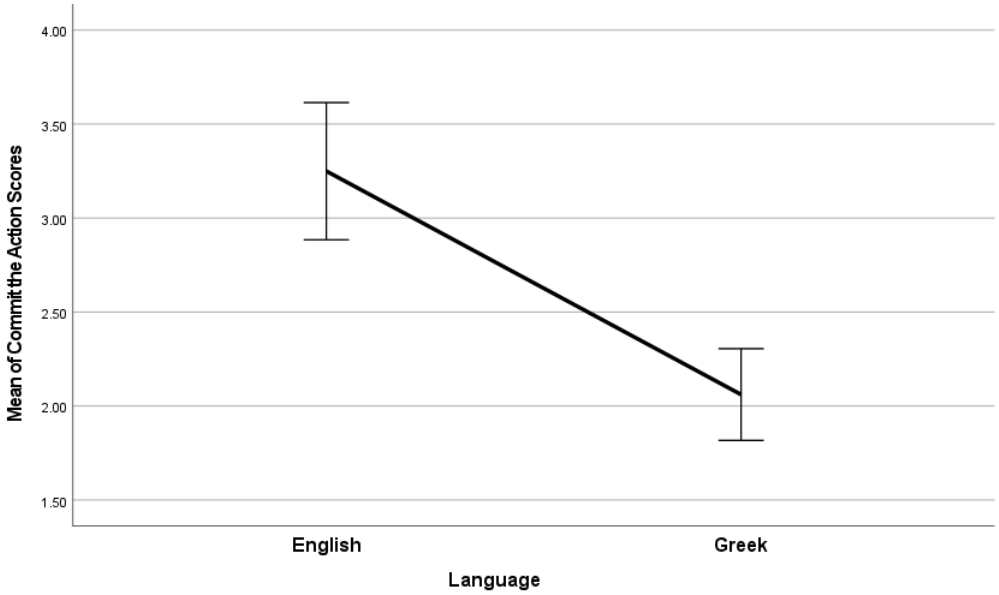
5.4 Results

5.4.1 FLE on Moral Judgment

Paired-samples t-test were conducted to compare the Commit the action scores between Greek (L1) and English (FL). There was a significant difference for committing the action scores in English (M=3.25, SD=1.89) and for committing the action scores in Greek (M=2.06, SD=1.26); $t(105)=-5.477$, $p< .001$. These results suggest that participants are more willing to commit the action in their FL English in contrast to their L1 Greek, see **Figure 20**.

Figure 20

Mean plots with 95% confidence intervals error bars of commit the action score between English and Greek.



5.4.2 FLE on Emotional States

In this part I conducted repeated-measures ANOVAs for each emotional scale, with two independent variables: language (Greek vs. English) and time (pre-test vs. post-test). The 2×2 design reflects the two levels of language (Greek, English) and the two levels of time (pre-test, post-test), examining how emotional responses change before and after making a moral decision. All effects are evaluated at a significance threshold of $p < .05$.

5.4.2.1 Negative affect

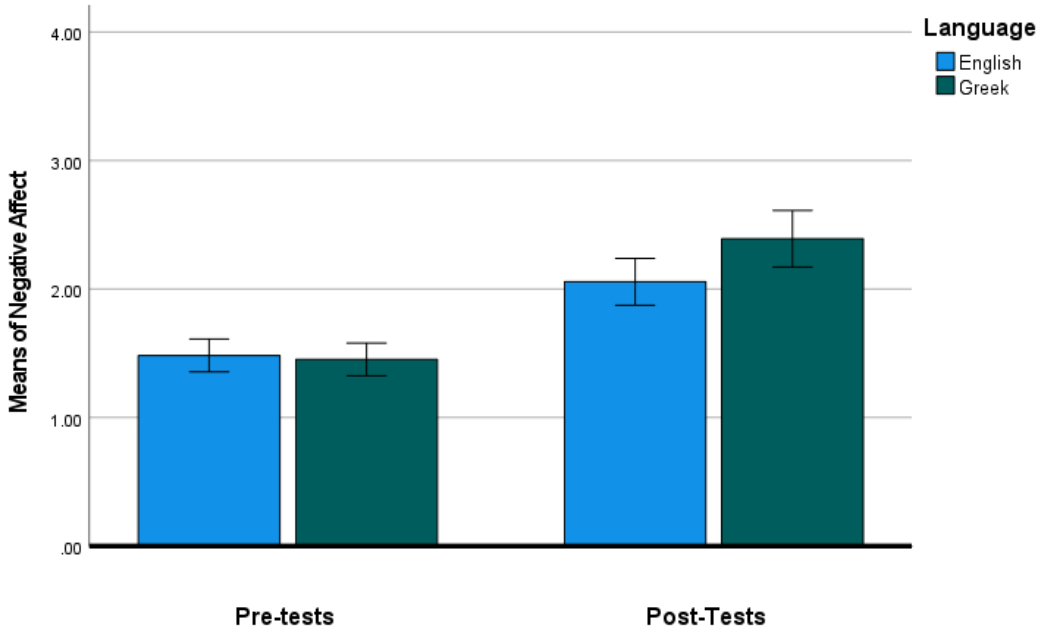
There was a significant main effect of the language on negative affect, $F(1, 105) = 14.473$, $p < .001$, $\eta^2 = .121$, where negative affect using the Greek language was significantly higher than using the English language. There was also a significant main effect of the time (pre-post) on negative affect, $F(1, 105) = 74.280$, $p < .001$, $\eta^2 = .414$, where negative affect in the post test was significantly higher than in the pre-test. There

was a significant interaction effect between the language and the time used, F(1, 105) = 14.949, p < .001, η^2 = .125. This indicates that the dilemma had different negative effects on people's scores depending on which language was used.

Planned comparisons showed no significant difference between languages in the pre-test (t (105) = 0.663, p = .509, d = 0.030, 95% CI [-0.061,0.121]) and significantly more negative affect on the post-test for the Greek language than the English (t (105) = -4.479, p < 001, d = -0.33491, 95% CI [-0.483,-0.187]), suggesting that although participants initially experienced similar affective responses across languages, exposure to the dilemmas in Greek led to a greater increase in negative affect compared to English. See **Figure 21**.

Figure 21

Mean plots with 95% confidence intervals error bars of means of negative affect between PANAS-X Pre-test and Post-tests.



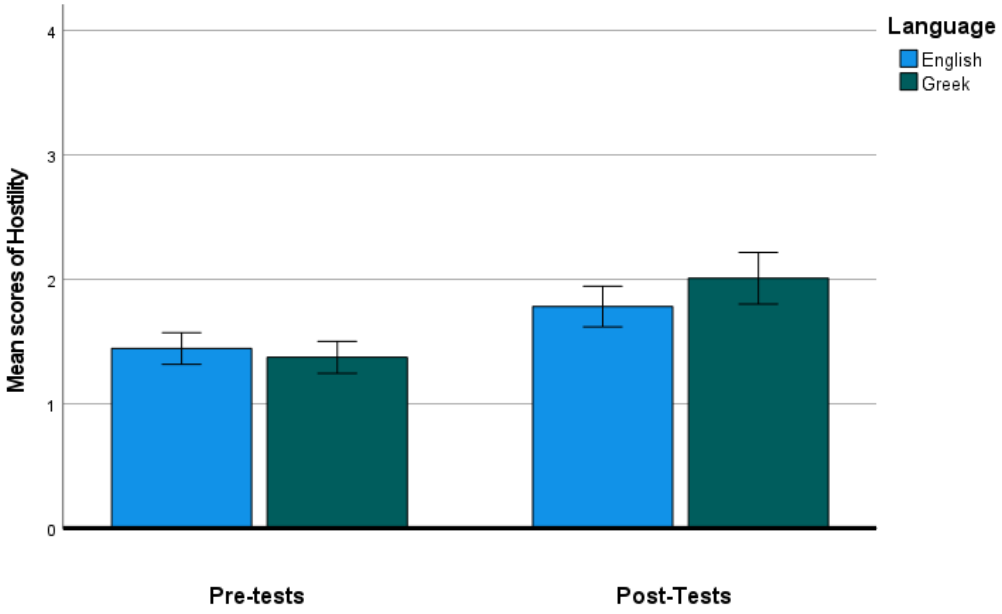
5.4.2.2 Emotions of hostility

There was no significant main effect of the language on hostility, $F(1, 105) = 2.949$, $p = .089$, $\eta^2=.027$ suggesting that participants' levels of hostility did not significantly differ depending on whether the dilemmas were presented in Greek or English. There was however, a significant main effect of the time (pre-post) on hostility, $F(1, 105) = 40.456$, $p < .001$, $\eta^2 = .278$, where hostility in the post-test was significantly higher than in the pre-test suggesting that exposure to the dilemmas led to a notable increase in participants' feelings of hostility regardless of the language in which they were presented. Furthermore, there was a significant interaction between the language and pre- to post- test use, $F(1, 105) = 10.818$, $p < .001$, $\eta^2 = .093$. This indicates that pre- and post- test use had different hostility on people's scores depending on the language used. In the Greek version the differences of emotions of hostility between the pre- and post- test increased significantly compared to the English version ($t(105)=3.489$, $p < .001$) indicating that the Greek language elicited a stronger emotional reaction, specifically a greater increase in hostility, than the English version of the dilemmas.

Planned comparisons showed no significant difference between languages in the pre-test ($t(105) =1.846$, $p =.068$, $d = 0.09$, 95% CI [-0.007,0.186]) and significantly lower hostility on the post-test for the English language than the Greek ($t(105) = -2.960$, $p = .004$, $d = -0.227$, 95% CI [-0.380,-0.075]), revealing that exposure to the dilemmas in Greek elicited a stronger hostile emotional response compared to English, despite no initial differences between the two languages. See **Figure 22**.

Figure 22

Mean plots with 95% confidence intervals error bars of mean scores of hostility between PANAS-X Pre-test and Post-tests.

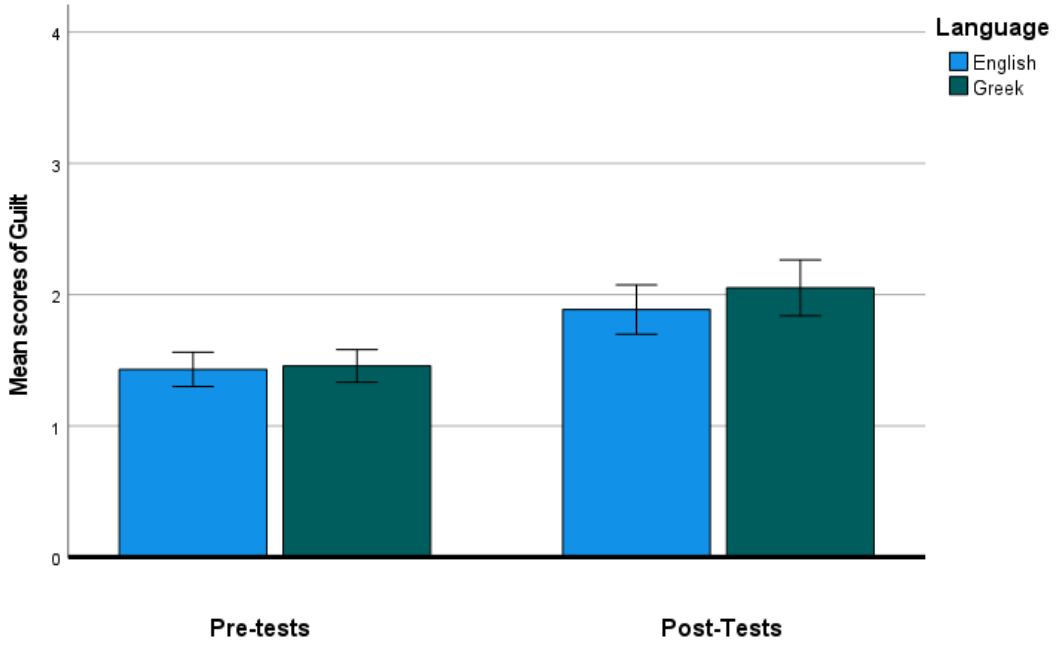


5.4.2.3 Emotions of guilt

There was a significant main effect of the language on guilt, $F(1, 105) = 4.902$, p .029, $\eta^2 = .045$, where emotions of guilt in the Greek language were significantly higher in contrast to the same emotions in the English language. There was also a significant main effect of the time (pre- to-post- test) on guilt, $F(1, 105) = 35.242$ $p < .001$, $\eta^2 = .251$, where guilt in the post test was significantly higher than in the pre-test. There was no significant interaction effect between the language and pre- or post- test, $F(1, 105) = 2.582$, $p = .111$, $\eta^2 = .024$, indicating that the change in hostility from pre- to post-test did not significantly differ between the Greek and English language conditions. See **Figure 23**.

Figure 23

Mean plots with 95% confidence intervals error bars of mean scores of guilt between PANAS-X Pre-test and Post-tests.

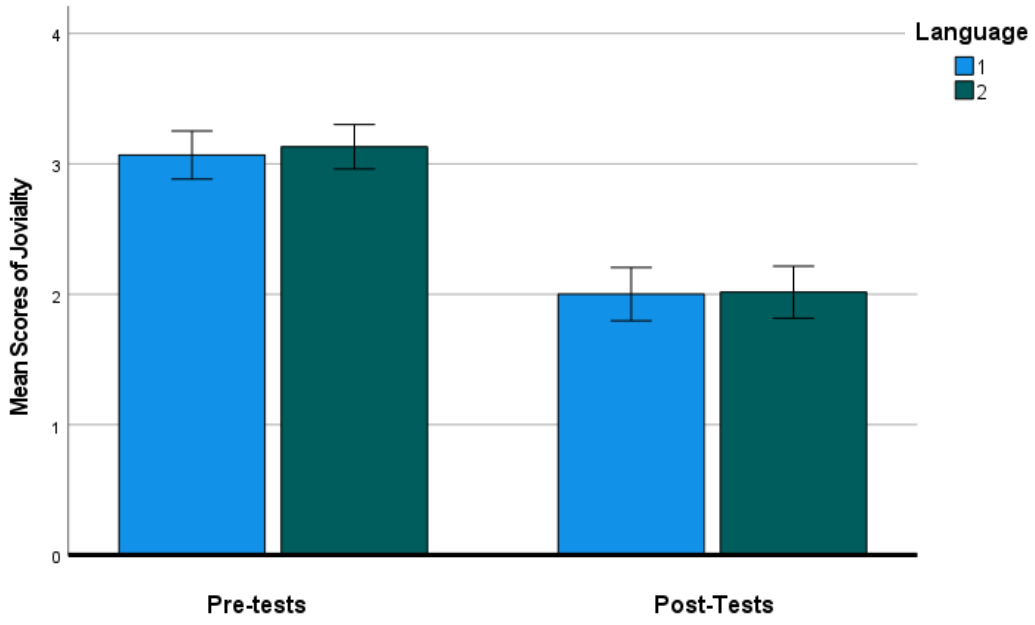


5.4.2.4 Joviality

There was no significant main effect of language on joviality, $F(1, 105) = 0.692$, $p = .407$, $\eta^2 = .007$ suggesting that participants' levels of joviality were not significantly influenced by whether the dilemmas were presented in Greek or English. However, there was a significant main effect of the time (pre- to post- test) on joviality, $F(1, 105) = 165.827$, $p < .001$, $\eta^2 = .612$, where joviality in the post-test was significantly lower than in the pre-test, indicating that exposure to the dilemmas substantially reduced participants' positive affect, regardless of the language in which they were presented. There was no significant interaction effect between the language and the time used on joviality, $F(1, 105) = 0.252$, $p = .617$, $\eta2 = .002$, suggesting that the decrease in joviality from pre- to post-test was similar across both the Greek and English language conditions. See **Figure 24**.

Figure 24

Mean plots with 95% confidence intervals error bars of mean scores of joviality between PANAS-X Pre-test and Post-tests.

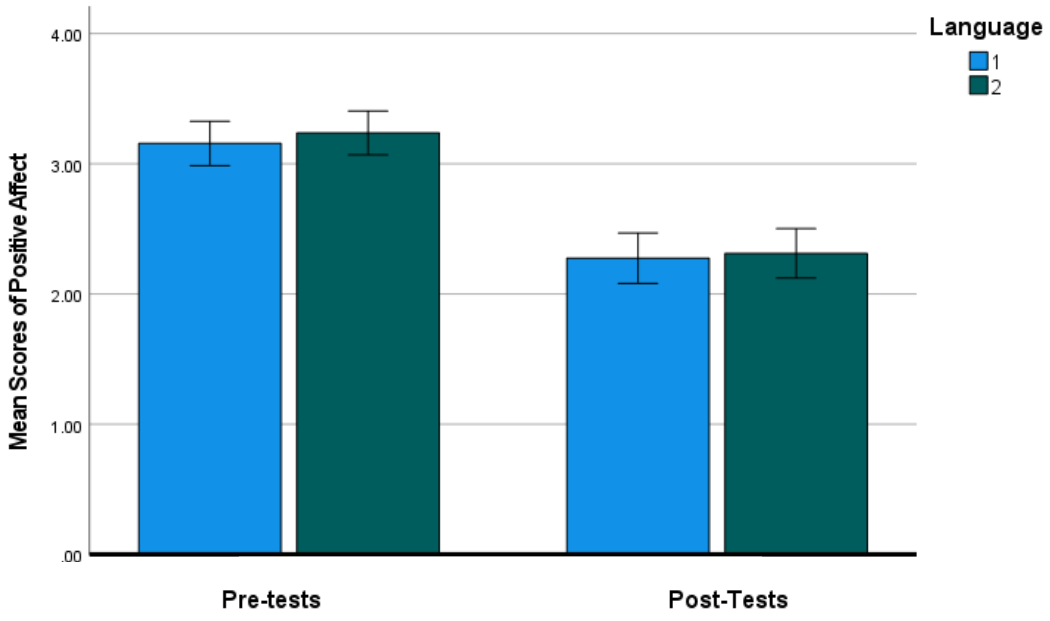


5.4.2.5 Positive Affect

There was no significant main effect of the language on positive affect, $F(1, 105) = 1.930$, $p = .168$, $\eta^2 = .018$ indicating that the language in which the dilemmas were presented did not significantly influence participants' overall levels of positive affect. There was a significant main effect of the time (pre-post) on positive affect, $F(1, 105) = 144.286$, $p < .001$, $\eta^2 = .579$, where positive affect in the post-test was significantly lower than in the pre-test suggesting that exposure to the dilemmas led to a substantial decrease in participants' positive emotions, regardless of the language used. There was no significant interaction effect between the language and the time used on positive affect, $F(1, 105) = 0.377$, $p=.540$ $\eta^2=.004$, meaning that the reduction in positive affect from pre- to post-test was consistent across both language conditions, with no significant differences in the pattern of change between Greek and English. See **Figure 25.**

Figure 25

Mean plots with 95% confidence intervals error bars of mean scores of Positive Affect between PANAS-X Pre-test and Post-tests.



5.4.3 FLE on Emotions Changed Score between Pre-Test and Post-Test

To further analyse the potential emotional effects across languages pre and post emotional responses were subtracted (using pre-test as baseline this were subtracted from post-test emotional estates) and paired-samples t-test were conducted to compare Emotional States between Greek (L1) and English (FL).

There was a significant difference in the paired samples scores between negative effect in English (FL) (M=0.57 SD=0.91) and negative effect in Greek (L1) (M=0,939 SD=1.13); t(105)= -3.866, p< .001 (**Figure 26**). The results suggest that there is more negative effect in emotions in the L1 rather than the L2. Also, there was a significant difference in the paired samples scores between emotions of hostility in English (FL) (M=0.34 SD=0.76) and emotions of hostility in Greek (L1) (M=0.64 SD=1.06); t(105)= -3.232, p= .002 (**Figure 27**). The results suggest that emotions of hostility are present more in the L1 rather than the L2.

However, there was a non-significant difference in the paired samples scores between positive effect in English (FL) (M=-0.88 SD=0.88) and positive effect in Greek (L1) (M=-0.93 SD=0.84); t(105)= 0.614, p=0.540. The results suggest that there is no significant difference between positive effect between the L1 and the FL. There was also a non-significant difference in the paired samples scores between Guilt emotions in English (FL) (M=0.46 SD=0.95) and positive effect in Greek (L1) (M=0.59SD=0.84); t(105)= -1.607, p = 0.111. The results suggest that there is no significant difference in emotions of Guilt between the L1 and the FL. Last, there was a non-significant difference in the paired samples scores between Emotions of Joviality in English (FL) (M=-1.07 SD=1.02) and Emotions of Joviality in Greek (L1) (M=-1.12 SD=0.99); t(105) = 0.502, p = 0.717. The results suggest that there is no significant difference between Emotions of Joviality between the Greek and English.

Figure 26

Mean plots with 95% confidence intervals error bars of negative affect scores between English and Greek

.

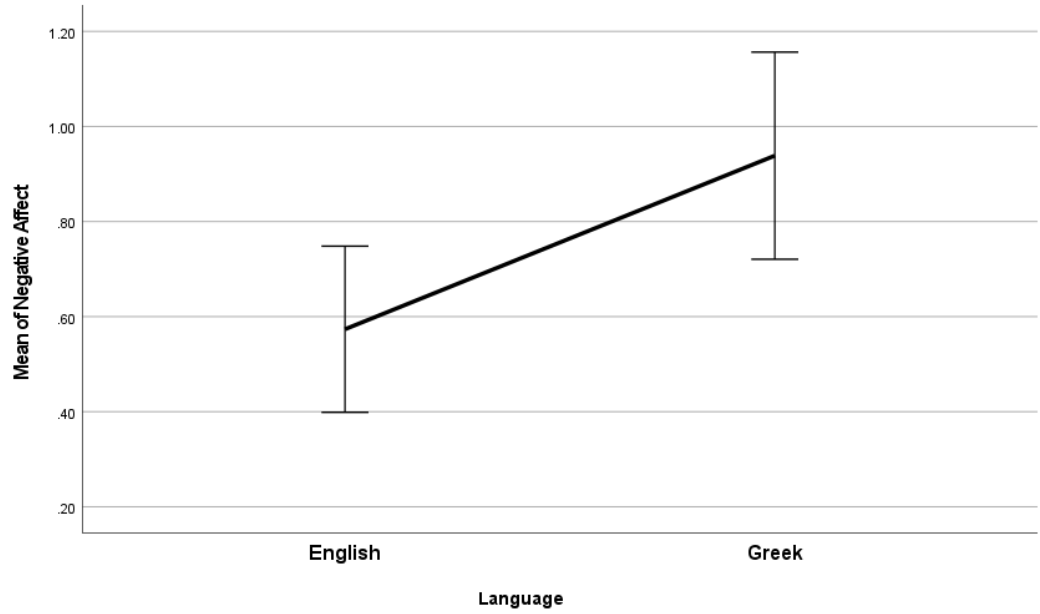
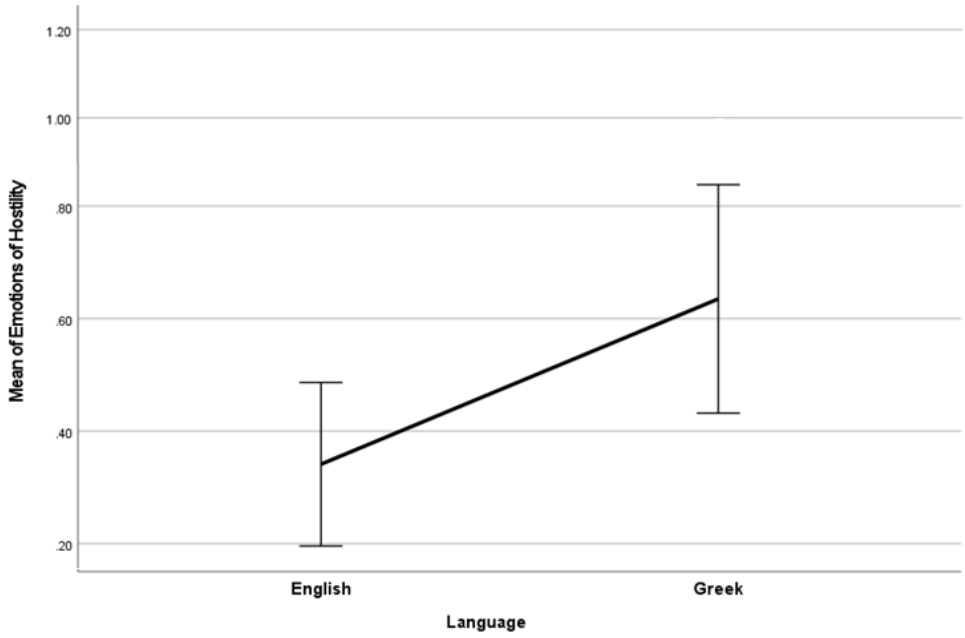


Figure 27

Mean plots with 95% confidence intervals error bars of hostility scores between English and Greek.



5.4.4 Validity of the within-subjects design (experiment 2)

To verify that participants were fully immersed in the designated linguistic setting and to validate the effectiveness of the language induction, a comprehension test was administered immediately following the language-inducing video. Participants who either failed to complete the test or answered incorrectly were excluded from further analysis. This method ensured that only individuals who effectively transitioned into the target language were included in the study, reducing the possibility of interference from other languages. The exclusion of these participants also served as a control measure, affirming the success of the language induction procedure.

As in experiment 1, participants were given one of four different versions of a questionnaire, each designed to ensure that no version included the same scenario in both their native and foreign language assessments. This counterbalanced approach aimed to minimize the influence of memory and repetition on participants' responses by effectively distributing systematic biases across conditions and reducing the impact of order effects (Brooks, 2012). However, due to the absence of comparable studies employing a within-subjects design in a single session to test the Foreign Language Effect (FLE), additional analyses were conducted to further verify the effectiveness of this experimental design.

To statistically examine whether the order in which participants encountered dilemmas (i.e., language block 1 vs. language block 2) influenced their moral judgments, independent samples tests were conducted for each language condition separately. These tests assessed whether repeating the task affected participants' moral decision-making.

The first analysis compared moral judgment scores for dilemmas encountered in English during the first block ($n = 54$) with those encountered during the second block ($n = 52$). Although the mean for the first block group ($M = 3.21$, $SD = 1.82$) was slightly lower than that for the second block group ($M = 3.29$, $SD = 1.98$), this difference was not statistically significant, $t(104) = -0.204$, $p = .839$. This finding suggests that the order in which dilemmas were presented in English did not significantly influence moral judgment.

The same analysis was conducted for moral judgment scores in Greek. This comparison involved dilemmas encountered in Greek during the first block (n = 54) and those encountered during the second block (n = 52). The mean for the first block group (M = 2.16, SD = 1.38) was slightly higher than that for the second block group (M = 1.96, SD = 1.15), but again, this difference was not statistically significant, t(104) = 0.794, p = .429. These results suggest that the order in which dilemmas appeared in Greek also did not significantly affect moral judgment.

Overall, these findings suggest that the within-subjects design effectively minimized the influence of prior exposure to similar tasks on participants' moral judgments. Furthermore, they indicate that the language-inducing task likely succeeded in establishing the intended linguistic mode, reducing the risk of biased responses due to task repetition. Thus, these results support the study's ability to reliably isolate the FLE in moral decision-making.

5.4.4.1 PANAS-X internal validity and reliability in the within-subjects design

One potential concern in within-subjects designs is the risk of carry-over or spill-over effects, where participants' responses in one condition may influence their responses in subsequent conditions (Greenwald, 2012). Initially, to minimize this risk, several methodological precautions were implemented. First, a language-inducing video with comprehension questions on a neutral topic that was included between the two language conditions to change the language mode served by also shifting participants' cognitive focus and reduce residual emotional influence. Research suggests that engaging in a cognitively demanding task, such as language processing, can facilitate emotional regulation and reduce the persistence of previously induced affective states (Hajcak et al., 2010). Second, participants encountered different moral dilemmas in their L1 and FL conditions, which helped mitigate the risk of habituation, desensitization, or direct repetition effects (Cushman et al., 2006). Furthermore, counterbalancing the order of language presentation across participants helped control for systematic sequence effects. However, as some residual emotional carry-over cannot be entirely ruled out the following tests were conducted for additional validation.

Therefore, for further validation paired-samples t-tests were conducted for each language separately to examine whether participants' pre-test PANAS-X scores differed significantly across conditions, which would indicate potential carryover effects. Additionally, a paired-samples t-test was conducted to assess whether pre-test PANAS-X scores in the English (EN) conditions differed significantly across the counterbalanced group that started with the English language context than those who did the English language context after doing the Greek language context. If there were carryover effects of repeated PANAS-X measurement, the latter group should have a different pattern than the former. The mean difference between pre-tests in English (EN), pre-Test EN 1 and Pre-Test EN 2 was 0.196 (SD = 2.025), with a 95% confidence interval ranging from -0.194 to 0.586, and the difference was not statistically significant, t(105) = 0.998, p = 0.321 (two-sided). Since the p-value exceeded the 0.05 threshold, the results suggest that participants entered each condition with comparable affective states, confirming that carryover effects were minimized. Therefore, any observed emotional changes during the tasks can be attributed to the experimental manipulations, rather than residual emotional effects from prior conditions. These findings enhance the study's internal validity and support the reliability of within-subject comparisons.

Additionally, a paired-samples t-test was conducted to assess whether pre-test PANAS-X scores in the Greek (GR) conditions differed significantly across the counterbalanced group that started with the Greek language context than those who did the Greek language context after doing the English language context. If there were carryover effects of repeated PANAS-X measurement, the latter group should have a different pattern than the former. The mean difference between Pre-Test GR 1 and Pre-Test GR 2 was -0.040 (SD = 1.109), with a 95% confidence interval ranging from -0.253 to 0.174. The difference was not statistically significant, t (105) = -0.368, p = 0.714 (two-sided). Since the p-value exceeds the 0.05 threshold, we fail to reject the null hypothesis, indicating that participants entered both Greek conditions with comparable affective states. The small mean difference and wide confidence interval further suggest that any observed differences are likely due to random variation rather

than systematic order effects. These results confirm that carryover effects were effectively minimized in this condition as no statistical differences were found for PANAS-X's pre-test scores participants who started the task with the Greek language context than those who did the Greek language context after the English language context., strengthening the study's internal validity and supporting the reliability of within-subject comparisons.

Furthermore, to assess the internal validity and reliability of the pre-test measures across participants in Greek and English, an Intraclass Correlation Coefficient (ICC) analysis was conducted using a two-way mixed-effects model. The results for the Greek Pre-Tests (including pre-test 1 and pre-test 2 for the two counterbalanced conditions, two different groups of participants) revealed an average ICC of 0.719 (95% CI: 0.588 – 0.809, $p < 0.001$), indicating good reliability across participants, with a consistent measurement structure. The single measures ICC was 0.562 (95% CI: 0.416 – 0.679, $p < 0.001$), suggesting moderate reliability at the individual level, with some variability in participants' responses. For Pre-tests in English the ICC results revealed an average ICC of 0.644 (95% CI: 0.477 – 0.757, $p < 0.001$), suggesting moderate-to-good reliability across participants. The single measures ICC was 0.475 (95% CI: 0.313 – 0.610, $p < 0.001$), indicating moderate consistency at the individual level with some degree of response variability (see supplementary **Figure S3** density plot). In addition, to the paired-samples t-tests presented earlier the significant ICC values confirm that the test maintains moderate-to-good internal validity, ensuring that the construct being measured remains relatively stable across participants.

5.5. Discussion

The findings of the current study demonstrate a robust FLE on moral judgement which is consistent with previous similar research (Chan et al. 2016; Cipolletti et al., 2016; Costa et al. 2014; Geipel et al., 2015; Christofi et al., 2023; Hayakawa et al. 2017; Wong & Ng 2018). This shows that people are more likely to make a utilitarian decision on a moral dilemma in their FL rather than their L1 (Hayakawa et al., 2016). It is believed that expressing emotions in the L2 is less anxiety-provoking, demonstrating

the emotional significance of one's L1 (Holmes & Mathews,2005). Therefore, in his study the aim was to shed light on the role of language emotionality and the exact emotions that could possibly drive the FLE on moral judgment by separating emotions into different categories, using the PANAS-X questionnaire (Horne et. al, 2016), in order to explore emotion categories individually and examine what emotions are induced and how the overall emotional state changes after participants have been exposed to the hypothetical moral dilemmas.

Predominantly, this experiment was conducted to compare Emotional States between the native and foreign language Greek (L1) and English (FL). The PANAS-X pre-to post analysis revealed the moral dilemmas successfully induced an overall change of emotions in both languages after the moral judgements on the dilemmas in both languages, but which emotional states have a significant difference between the two languages and therefore the FLE?

When emotionality is measured using rigorous measures (PANAS-X) and by diving emotions into different categories the results revealed that there was a significant difference between negative effect scores in English (FL) and negative effect scores in Greek (L1) suggesting that there is more negative effect in emotions in the L1 rather than the L2. Also, there was a significant difference in the paired samples scores between emotions of Hostility in English (FL) and emotions of Hostility in Greek (L1). The results suggest that emotions of Hostility are present more in the L1 rather than the L2. Whereas, positive, joviality and guilt emotional states even though they change from pre- to- post-tests it is regardless of language, therefore no FLE.

This study is consistent with the emotion-reducing hypothesis where foreign language messages usually elicit less significant emotional responses in contrast to the L1 (Dewaele, 2004; Harris, 2004; Harris et al., 2003; Hsu et al., 2015; Iacozza et al., 2017; Zheng et al., 2020) but the results of the current experiment profoundly shed light on the possibility that Negative Affect emotions could actually be a factor that mediates or moderates the FLE on moral judgment. The results are in accordance with Dylman and Bjärtå (2019) who predominantly investigated the role of the use of negative emotions and its effect of L2 use, where they found a clear effect of L2 use on negative

emotions. Specifically, responding in Swedish (L1) led to higher distress ratings, whereas responding in English (L2) reduced distress.

The reduction of negative feelings and emotions of hostility in the FL, as found in this PhD, could also be related to research where FLE is found on all crime-related moral dilemmas as the fact that the FL diminishes crime severity judgment could derive by this phenomenon (Woumans et al.,2020). Additionally, the results could also be in accordance with research on risk judgment and the FLE where reduction on negative emotions on judgment of risk were found (Hadjichristidis, et al., 2015) and Affective valence on less negative feelings towards bad-luck scenarios in the FL (Hadjichristidis, et al., 2019).

This could also be linked to research that suggests that processing emotions in a non-native language can reduce their intensity. This effect, documented by Dewaele (2010) and Pavlenko (2012), indicates that negative emotions may be less potent when experienced in a foreign language. Which implies that individuals might feel less intense negative emotions when interpreting information in a foreign language, as the emotional impact is frequently lessened in contrast to their L1 (Harris et al., 2003).

Furthermore, the results could be related to the study by García-Palacios et.al., (2018) who investigated the role of fear, which falls into the category of negative affect, where fear conditioning demonstrated less emotional reactivity in the FL in contrast to the L1. Similarly, it seems that native language conditions involving negatively valenced imagery resulting in greater increase of anxiety with no significant changes in other emotions (Holmes & Mathews, 2005). Considering the likelihood that participants might have been mentally visualizing each dilemma, wherein they were tasked with making utilitarian decisions, particularly those scenarios where they were required to save themselves alongside others, could have potentially heightened anxiety levels. Consequently, this could have led to a more pronounced negative emotional impact on their native language (L1)

Chapter 6. General discussion

The present PhD thesis investigated the possible factors that drive the FLE on moral judgement, one of the most cutting-edge topics in the field of bilingualism: Does foreign language use influence cognitive processing, resulting in more rational and utilitarian decision-making?

The theoretical framework of the FLE on moral judgment refers to the phenomenon whereby speakers using an FL tend to make more utility-driven decisions, such as those that prioritise the greater good (Greene, 2007), compared to decisions made in their native language. Moral judgment is typically assessed using hypothetical moral dilemmas (Costa et al., 2014; Hayakawa et al., 2017), in which individuals are asked to make imagined ethical choices, such as sacrificing one person to save, for example, five or more others. For instance, one scenario may involve pushing a hurt person through the debris of a burning building in order to escape along with five other people (Wong & Ng, 2018).

The experiments in this thesis revealed a clear effect of the FL on moral judgment, consistent with prior literature (Chan et al., 2016; Cipolletti et al., 2016; Costa et al., 2014; Geipel et al., 2015; Hayakawa et al., 2017; Stankovic et al., 2022; Wong & Ng, 2018). Based on the abovementioned literature one would expect that presenting moral dilemmas in an FL would lead to more utilitarian decisions (which means that people are more willing to take one life to save five to ten other individuals) in their FL but not in their L1. The current thesis comprehensively supports the FLE phenomenon, contributing to the limited research involving Greek (L1)-English (FL) bilinguals in the existing literature, through the novel inclusion of this new group not tested in the exact manner before.

But what drives this intriguing phenomenon? The theoretical framework reports two possible main factors that potentially mediate or moderate the FLE: The cognitive load hypothesis and the emotion-reducing hypothesis. This PhD thesis was designed in a way to shed light into the uncertainty between the two main proclaimed factors and

weigh evidence towards the one or the other or both on what drives the FLE 1: cognitive load; language disfluency that results from lower proficiency in an FL or 2: reduced FL emotionality due to the FL distance in contrast to the native language (Hayakawa et al., 2016).

The cognitive load account indicates that the FLE is influenced by factors such as language proficiency; higher foreign language proficiency was hypothesised to modulate the degree to which language affects moral judgments, potentially diminishing the FLE (Costa et al., 2014). The emotion-reducing hypothesis refers to the phenomenon that emotional responses in an FL induce less emotional responses in contrast to the L1 (Iacozza et al., 2017) and create a detachment and that our L1 is considered our most emotional one (Geipel et. al., 2015). The prominent division between these two central theories cognitive load hypothesis and emotion reduced hypothesis (extensively discussed in the theoretical background chapter 2) and the existing ambiguity on which factor drives the FLE (Circi et al., 2021; Del Maschio et al., 2022; Dylman & Bjärtå, 2019), uncovered the need that these possible factors should be independently, distinctly and thoroughly examined.

However, could the reason for this uncertainty between the two main factors that motivate the FLE be the dense reliance on self-measures? (Circi et al., 2021; Privitera et al., 2023) e.g., self-rated FL proficiency (Costa et al., 2014; Wong & Ng, 2018), or by eliciting language emotionality using single emotion-rating questions; e.g., rate how upset this dilemma makes you feel (Geipel et al., 2015; Wong & Ng, 2018). Consequently, using explicit measures, instead of relying on self-measured FL proficiency, fluency, and emotionality, for both theories, was the approach this PhD employed to investigate each factor's role on the FLE. Hence, to address the above, the novelty of this thesis is that I fully examine both factors using rigorous standard measures in all experiments (cognitive load hypothesis: 1: proficiency and FL fluency, emotion-reduced hypothesis: 2: FL emotionality), in order to provide concrete evidence toward one factor, the other, or both.

Another profound element of the current PhD is the design in all experiments. In contrast to the majority of previous research a within-subjects design (also found in Mills & Nicoladis's, 2020) was implemented (comparing responses from the same

participants in both languages) instead of the widely used between-subjects design (that compared responses from two separate groups of participants: one group responds only in their L1 and the other one only in their FL, (Costa et al., 2014; Geipel et al., 2015; Wong & Ng, 2018).This design was implemented explicitly for the reason that moral judgment is such a personal insight for each individual therefore, moral resonance could vary greatly from one person to the other. This has been described by Reynolds & Ceranic (2007) as the moral identity which is a type of identity centred on the moral aspects of a person's self-concept (Bergman, 2004). It functions as a self-governing mechanism, guiding behaviour and motivating moral judgement for each individual. Moral identity in combination with social consensus; the degree of social agreement on whether an act is considered right or wrong (Jones, 1991); generates an individual's moral judgment (Reynolds & Ceranic, 2007). Hence, as moral judgment is idiosyncratic, using a within-subjects design provides a clearer image, with more increased statistical power as it focuses on the conditions variances and the experimental control, excluding probable individual discrepancies of moral rules and thinking style patterns between the participants tested that could derive from the between-subjects design (Białek & Fugelsang, 2019). Consequently, this may prompt an inquiry into the suitability of employing a between-subjects design for this kind of experiments (involving moral judgment), particularly in the context of evaluating emotionality. Such considerations are critical for determining the methodological rigor and validity of the findings related to emotional assessments. By implementing the aforementioned design, the results of the main factors are revealed as follows.

6.1 Cognitive load hypothesis

The second aim of this study was to examine the role of FL proficiency and fluency on moral decisions in the FL aka FLE. Based on the theoretical framework above supporting the cognitive load hypothesis (Costa et al., 2014; Geipel et al., 2015; Hayakawa et al., 2016; Wong & Ng, 2018), in experiment 1, I predicted that the lower the FL proficiency or fluency the more FLE would be present in contrast to more proficient and more fluent bilinguals. Moral dilemmas (hypothetical scenarios that ask participants to hypothetically kill an individual in order to save 5-10 people from certain death; see Appendix 3a.) were used in all experiments (Wong & Ng, 2018).

Experiment 1 tested FL proficiency by using a standard proficiency test (Oxford QPT) to measure FL proficiency in Greek-English bilinguals and FL fluency by using four standard fluency tests (two letter and two category standard fluency tests) instead of only relying on self-rated proficiency which the majority of previous research used (Costa et al., 2014; Wong & Ng, 2018).

When using rigorous standard proficiency measures, the current study found no significant correlation or effect between FL proficiency and the FLE, which means that regardless of the participants high or low proficiency there was no difference in their willingness to make a utilitarian decision on the hypothetical moral dilemmas they were presented with. These findings are inconsistent with research that supports the cognitive load account (see Costa et al., 2014; Hayakawa et al., 2017; Wong & Ng, 2018) but in accordance with Miozzo et al. (2020) who argued that the FLE does not arise due to cognitive resource demands in high proficient participants. In addition, experiment 1 of this PhD confirmed Miozzo et al.'s (2020) findings, expanding them to both high and low proficiency groups instead of focusing merely on high-proficient participants.

A possible explanation of this deviation between previous results could be the use of mostly self-rated proficiency calculations in the majority of previous research (see Circi et al., 2021; Del Maschio et al., 2021). Therefore, the urge for the need to implement standardised measures for FL proficiency in the FLE context (Anton et al., 2020; Circi et al., 2021; Marian et al., 2007; Privitera et al., 2023) was undoubtedly a great calling in tackling down the existing ambiguity on the role the cognitive load (FL proficiency) plays on the FLE by rigorously shedding light into this proclaimed factor. Another explanation could be, as abovementioned, the probable interference between people with variances in the between-subjects design previously used (Białek & Fugelsang, 2019).

In experiment 1, on foreign language fluency, the consistent FLE was merely pointed in the high fluency group. These outcomes contradicted the initial expectations where the FLE was expected to be more prominent in the low fluency group of bilinguals rather that the higher fluency group. These variations could also be compelled by the possibility that the higher fluency group coincidentally could be overall less

likely to commit the action in their L1 too (Tassy et al.,2013). While the language context can influence how people make moral decisions, some individuals have a natural tendency towards utilitarianism, which is significantly influenced by their cognitive abilities and personality characteristics (Tassy et al., 2013; Del Maschio et al., 2022). In this case having employed the within-subjects design enabled the cross-check between fluency and comprehension scores as a sequence associated to the FLE, where a clear linear trend cannot be observed (see Chapter 4: Figure 18A and 18B). This further extends the point about the use of the within-subjects design and the confounding role of proficiency.

My findings showed no significant effect of L2 proficiency or fluency on the FLE, which suggests that cognitive load alone may not fully explain the observed differences in moral decision-making. If increased cognitive load were the primary mechanism behind the FLE, we would expect greater L2 proficiency to reduce or even eliminate the effect, as more proficient speakers experience lower cognitive demands when processing information in L2. However, it is acknowledged that L2 proficiency in our sample was lower than L1 proficiency, even for highly proficient speakers. It is possible that even the most proficient L2 users experience some degree of additional cognitive load compared to their native language, which our study may not have been sensitive enough to detect. Despite these considerations, prior studies (e.g., Costa et al., 2014; Cipolletti et al., 2016) have demonstrated FLE effects even in bilinguals with near-native L2 proficiency, suggesting that factors beyond cognitive load—such as emotional distance or reduced intuitive processing—play a role. Future research could address this limitation by including a wider range of L2 proficiency levels or by directly manipulating cognitive load to isolate its impact on moral decision-making.

As far as it concerns the type of the moral dilemmas, self-preservation (SP): the utilitarian choice will save themselves along with others) vs non self-preservation (NSP); the utilitarian choice only involves other people; (Mills & Nicoladis, 2020) the results of this PhD were consistent with studies that supported that individuals are more inclined towards making a utilitarian choice when they were hypothetically at risk of losing their life too (SP) in contrast to dilemmas that merely concerned other individuals which could be due to the possible negative emotions that could be induced from moral

dilemmas that involve saving someone's self (Mills & Nicoladis, 2020; Sachdeva et al., 2015). However, these results cannot be related to the FLE as these choices were regardless of the language used however this indicated the need to investigate FL emotionality in more depth. Therefore, experiment 2 extensively investigated foreign language emotionality by looking into different emotional states on the FLE separately.

6.2 Reduced Emotionality

Literature has demonstrated that there is an emotional reduction in the FL as emotional words are not as profoundly encoded (Dewaele, 2008) as in our native language. That could derive from the potential psychological detachment that occurs in a foreign language (Pavlenko & Dewaele, 2002). In contrast, in the native language there is an innate familiarity to emotional words due to the frequency of use (in different contexts) and the deep level they are mentally stored (Altarriba, 2003).

The emotion reduced account discusses how foreign language use is less emotional due to the lack of intensity on mental images as they are reduced with FL use (Hayakawa & Keysar, 2018) and that messages conveyed in a foreign language typically evoke weaker emotional reactions compared to those delivered in one's native language (Hsu et al., 2015; Iacozza et al., 2017; Zheng et al., 2020). This difference arises because foreign languages are often learned in environments that lack emotional depth, unlike native languages, which are acquired in rich, emotionally charged contexts (Cipolletti et al., 2016). As a result, individuals process and respond to foreign language communications with less emotional intensity.

Most studies on the emotion reduced hypothesis measure language emotionality by asking participants to rate how upset they feel after reading each dilemma (Wong & Ng., 2018) which is also what I included in my first experiment. When eliciting language emotionality using a single question in experiment 1 (participants had to rate how upset the dilemma made them feel) language emotionality had no significant main effect on language which means that emotionality scores in the FL where no different in the native language which means there is no FLE observed.

However, the use of robust emotional measures; pre-test and post-test PANAS-X in experiment 2 accelerated the emotional states before and after participants were exposed to the hypothetical moral dilemmas by measuring emotions in each category and providing the opportunity to shed light into the role of the different categories respectively (negative affect, positive affect, hostility, guilt and joviality; Horne & Powell, 2016) and therefore the emotion reduced hypothesis on the FLE.

In experiment 2, moral dilemmas were once again employed to assess participants' moral judgments. The use of moral dilemmas however, in measuring moral judgment, induces negative emotions, as hypothetically having to kill a person could produce negative effect emotions such as fear, sadness, anxiety, sadness hostility and so on (Horne & Powell, 2016). This could be because participants' moral principles are conflicted, and the potential costs of their choices induce negative emotions (Tangney et al., 2007).

The results of the second experiment revealed a non-significant difference between English (FL) and in Greek (L1) in the emotional categories of positive effect, Guilt and Emotions of Joviality. However, there was a significant difference in the paired samples scores between the emotional category of negative effect in English and negative effect in Greek between emotions of Hostility in English which were more present in the native language rather than the FL of participants. This could be explained by the research that indicates that using a foreign language can significantly alter emotional responses, particularly by diminishing the intensity of negative emotions. For example, Caldwell-Harris (2015) and Pavlenko (2017) have found that the psychological distance created by using a foreign language, results in weaker emotional reactions, such as fear, anger, or sadness. Consequently, decisions made in a foreign language context are less influenced by immediate emotional discomfort and more by rational deliberation. On the other hand, positive emotions are known to be consistently expressed across languages, without exhibiting as great intensity as negative emotions in one's native language due to several factors. Universal experiences and expressions of positive emotions, such as happiness and joy, are similar across cultures (Ekman, 1992). In contrast, cultural norms and socialization practices influence the intensity of negative emotions (Kitayama & Markus, 1994). Emotional memories,

especially negative ones, are more vividly recalled and expressed in one's native language (Pavlenko, 2005). A native language's richer and more nuanced emotional vocabulary also contributes to the stronger expression of negative emotions (Wierzbicka, 1999). Therefore, this could explain why negative emotions, elicited by dilemmas, contribute to the emergence of the FLE, while positive emotions appear to have no influence.

Additionally, my findings align with those of Dylman & Bjärtå (2019), who demonstrated that individuals experience lower emotional distress when processing negative content in the L2 compared to their L1. Their study found that participants who read distressing texts in their native language (Swedish) reported reduced distress when they later answered questions in L2 (English), whereas those who processed everything in L1 exhibited stronger emotional responses. This suggests that using an L2 may create emotional distance, weakening the impact of distressing or emotionally charged content. In the context of my study, this phenomenon may help explain why participants exhibited different emotional responses depending on the language used. My study involved moral dilemmas which required participants to make emotionally charged decisions therefore while processing these dilemmas in the L2 a diminished emotional connection to the content may have occurred, influencing their choices. This aligns with research suggesting that emotions are more deeply rooted in one's L1, while L2 often feels more detached (Keysar et al., 2012), leading to a more emotionally neutral processing of distressing content. Dylman & Bjärtå's (2019) findings highlight that, participants experienced stronger emotional reactions when both reading and responding to negative texts in their native language. Conversely, when they read in L1 but processed the content in L2, distress levels were reduced. This shift in emotional intensity suggests that using a second language creates a form of emotional distancing, where the emotional weight of distressing stimuli is lessened when processed in the L2. In my study, a similar mechanism may have been at play—participants engaging with dilemmas in L2 might have found the emotional stakes to be lower, resulting in responses that were less emotionally driven. Specifically, negative emotions such as sadness may originate from the unavoidable harm or loss that comes with making tough choices (e.g. such as those in moral dilemmas used in my study), particularly when there are no obvious solutions (Haidt, 2003) influencing moral judgments (Koenigs &

Tranel, 2007).

The findings of this study align with and extend the emotion reduced hypothesis, as demonstrated by Kyriakou et al.,(2023), who found that bilinguals using their L2 exhibited lower emotional arousal and a greater propensity for utilitarian reasoning in moral dilemmas. Their study revealed that moral decisions in L2 were associated with a reduction in high-arousal emotional words, suggesting that diminished emotional intensity facilitates rational, utilitarian decision-making. Similarly, the results of experiment 2 in this study demonstrated that while overall emotionality ratings did not differ significantly between L1 and L2, specific negative emotional categories—such as hostility and negative affect—were significantly less pronounced in the FL. This finding reinforces the argument that emotional blunting, rather than cognitive load, underlies the FLE. Furthermore, the absence of a significant correlation between FL proficiency and moral decision-making challenges cognitive load accounts (Costa et al., 2014; Hayakawa et al., 2016) and suggests that the reduction in emotional resonance in L2 is independent of linguistic competence. The current study contributes to this debate by employing rigorous standardized fluency and proficiency measures, unlike previous research that predominantly relied on self-rated assessments. Additionally, the findings suggest that the differential impact of emotional categories on FLE warrants further investigation, particularly in understanding why negative emotions appear to drive the FLE more than positive emotions. This supports previous arguments (Caldwell-Harris, 2015; Pavlenko, 2017) that the psychological detachment in L2 weakens negative emotional responses, thereby increasing utilitarian tendencies in moral decision-making.

It needs to be stressed that this study is among the first to explicitly recognize the distinction between SP and NSP dilemmas within the framework of the FLE. Previous research has primarily focused on moral decision-making in utilitarian versus deontological terms, often overlooking the extent to which the decision-maker's personal stake influences their ethical judgments (Christensen et al., 2014; Greene et al., 2001). By differentiating between SP and NSP scenarios, this study highlights how individuals are significantly more inclined to endorse a utilitarian decision when their own survival is at risk, as opposed to dilemmas where only third-party individuals are affected. This distinction is crucial, as what seems like utilitarian reasoning may

actually stem from self-preservation rather than a true concern for collective welfare. Research shows that people are much less likely to condone harm when they themselves benefit, highlighting how moral judgments are shaped by personal involvement (Miller & Cushman, 2013). These findings call for further investigation into whether such decisions are retroactively framed as utilitarian or reflect a separate psychological process.

To sum up, this PhD has contributed to the FLE phenomenon by extending research on a group of informants, Greek-English bilinguals, never tested before explicitly on the FLE on moral judgment. With the rigorous proficiency and fluency tests and by employing the within-subjects design that eliminates external factors (Mills & Nicoladis (2020) keeping individual differences, such as age, gender, intelligence, etc. consistent the results of the first 2 experiments potentially could override the relation between the FLE and the cognitive load hypothesis in this study. However, the pivotal finding of this PhD was revealed in the second experiment which profoundly supports the emotion-reducing hypothesis specifically with negative effect emotions being the factor that mediate or moderate the FLE on moral judgment.

6.3 Limitations of the study

A potential limitation of the study pertains to the wide range of proficiency levels present in experiment 1. Although this issue was mitigated by focusing on contrasting proficiency cases within a group of 50 participants (25 highest scores and the 25 lowest scores), future research could enhance the robustness of the findings by employing a larger sample size of only contrasting proficiency groups. Specifically, future research should concentrate exclusively on participants with very high proficiency levels (C1-C2) and very low proficiency levels (A1-A2) according to the CEFR scale, thereby excluding the intermediate proficiency levels (B1-B2).

Another point that needs to be considered is that, in the first experiment, I did not find a significant effect of language context on emotionality ratings when participants were asked how upset they felt while reading the dilemma. This contrasts with previous research that employed between-subjects designs, which have reported a FLE in emotionality. One possible explanation is that my within-subjects design

effectively controlled for individual differences, thereby reducing variability that may have contributed to the effects observed in between-subjects studies. Alternatively, it could be argued that the within-subjects approach introduced carry-over effects, potentially influencing participants' responses in the second language condition. To mitigate language mixing, I incorporated a language mode induction procedure, including comprehension questions, in both experiments. Importantly, I tested for block effects (in both experiments) and found no significant differences, indicating that order effects did not systematically affect the results. This suggests that carry-over effects were unlikely to have played a substantial role in either experiment. However, in experiment 2, where I examined changes in how upset participants felt before and after completing the moral dilemma task, I did observe a significant effect of language context. This suggests that, while language may not have influenced the immediate emotional response to each dilemma in experiment 1, it did contribute to the overall emotional experience over time in experiment 2. This finding underscores the need to consider how language effects evolve throughout an experiment, rather than focusing solely on momentary emotionality ratings. Future research could further explore whether language influences emotional processing differently depending on whether it is measured before and after a moral dilemma task or at a single moment in time. Additionally, alternative experimental designs could be explored to enhance language separation while maintaining the advantages of a within-subjects methodology.

Last, concerns about carry-over effects in moral dilemma experiments assume that prior decisions will necessarily shape subsequent choices in a deterministic manner. In this PhD block effects analysis and statistical validation were carried out in sections 4.6.2 and 5.4.4 as a verification of the within-subjects design used. However, research suggests that moral decision-making is largely context-dependent and guided by distinct cognitive and emotional processes rather than rigid consistency (Greene et al., 2001). While repeated exposure to similar dilemmas may lead to response patterns, empirical studies indicate that individuals do not automatically generalize their choices across different scenarios; instead, they evaluate each dilemma based on situational factors, emotional engagement, and cognitive control mechanisms (Christensen & Gomila, 2012). The dual-process theory of moral judgment posits that deontological and utilitarian reasoning operate through different neural and psychological pathways,

meaning that making a utilitarian choice in one scenario (e.g., pushing a man off a bridge) does not necessitate a utilitarian response in another (e.g., shooting an injured diver) (Greene, 2014).

Additionally, well-designed moral dilemma studies often employ counterbalancing techniques to mitigate order effects, ensuring that choices remain independent rather than being influenced by prior responses (Brooks, 2012). Thus, while the risk of carry-over effects exists, it is neither an inevitable flaw nor an uncontrollable factor in experimental setups designed to examine moral reasoning. Another important consideration is that previous research using both personal and impersonal dilemmas (e.g., Wong & Ng, 2018) has found that the FLE does not appear in impersonal dilemmas and in this study certain types of dilemmas, such as those that do not involve self-preservation were rated as less emotional and induced a lower commit the action scores. Therefore, these results demonstrate a focus on evaluating each dilemma based on its individual factors, especially since negative effect could be what mediates or moderates the FLE on moral judgment (experiment 2).

To conclude, this PhD dissertation was meticulously structured to assess the role of the two proposed as the predominant hypotheses regarding the factors influencing the FLE on moral judgement; the emotion-reducing hypothesis and the cognitive load hypothesis (Hayakawa et al., 2017). This PhD dissertation attempted to clarify the confusion in the literature regarding the role of two abovementioned hypotheses and their potential interconnection by employing rigorous measures. The results potentially overrule the cognitive load hypothesis and support the emotion-reducing hypothesis as the factor that influences or regulates the FLE on moral judgement.

Chapter 7. Conclusions

In conclusion, all experiments in the present PhD research are consistent with previous research with a robust FLE on moral judgement (Chan et al., 2016; Cipolletti et al., 2016; Costa et al., 2014; Geipel et al., 2015; Hayakawa et al., 2016; Hayakawa et al., 2017; Stankovic et al., 2022; Wong & Ng, 2018).

Rigorous measures of proficiency and fluency scores were used in combination with a within-subjects design that eliminated external factors that could arise by comparing different groups of people (Białek & Fugelsang, 2019). But with this model neither proficiency nor fluency were factors that mediated or moderated this effect in the first experiment. This important finding could possibly override the assumption that the FLE occurs due to low language proficiency (Costa et al., 2014) and therefore challenge the role of the cognitive load theory on the FLE.

Moreover, the actual proficiency scores were to some extent inconsistent with the self-rated comprehension scores of each dilemma demonstrating once again that self-rating scores could vary to a certain extent from the reality (Mauss & Robinson, 2009) creating a big question on extensive previous research that only used this method on the FLE (Privitera et al., 2023; see also self-rating proficiency by Tomoschuk et al., 2019).

In the first experiment language emotionality was elicited using a single question on how the moral dilemmas made participants feel (Wong & Ng.,2018). As predicted SP dilemmas induced more utilitarian responses in contrast to NSP ones and were perceived as more emotional probably due to the hypothetical danger the agent was involved in (Hoffman, 2000). However, this did not have an effect on the FL but what seemed to play an important role in moral judgement, regardless of language, however, was the type of dilemma. This study is among the first to incorporate moral dilemmas into the FLE framework while explicitly acknowledging the distinction between SP and NSP dilemmas. Prior research has largely framed moral decision-making in terms of utilitarianism versus deontology, often overlooking how an

individual's personal stake in a dilemma shapes their ethical choices (Christensen et al., 2014; Greene et al., 2001). By emphasizing the role of SP and NSP dilemmas, this study demonstrates that individuals are significantly more likely to endorse a utilitarian course of action when their own survival is at stake compared to situations affecting only third parties. This differentiation is critical, as it suggests that what appears to be utilitarian reasoning may, in some cases, stem from self-preservation instincts rather than a deliberate commitment to maximizing overall welfare. Moreover, empirical evidence reinforces that individuals are far less willing to justify harm when they themselves stand to benefit, aligning with previous findings that moral permissibility judgments are influenced by whether the decision-maker is directly involved in the outcome (Miller & Cushman, 2013). Given these insights, future research should explore whether SP concerns are rationalized retrospectively as utilitarian reasoning or if they constitute a separate psychological process. Additionally, further investigation is needed to determine whether self-interest has the capacity to override initial moral hesitation specifically on the FLE. Indeed experiment 2 profoundly demonstrated the role of Negative Affect emotions that could indeed be a contributing factor towards the FLE on moral judgement linking this back to the theoretical background chapter (see section 2.2) where our native language is more emotional (Pavlenko, 2005; Pavlenko, 2006; Dylman and Bjärtå, 2019). In conclusion, the experiments of this study support the emotion- reducing hypothesis account with the category of negative effect emotions being a factor that roots the FLE and on the other hand, override the cognitive load hypothesis account on the FLE when rigorous proficiency and fluency measures were employed.

While this study provided novel insights into the FLE by demonstrating that FL proficiency might not seem to be mediating the effect, but instead negative emotions seem to do, several important questions remain for future research. One key methodological advancement of this study was the use of a within-subjects design, which allowed for greater control over individual differences and a more precise measurement of the FLE. Given that most prior research has relied on between-subjects designs (Costa et al., 2014; Hayakawa et al., 2017), future studies should adopt within-subjects methodologies to enhance reliability and reduce potential confounds related to

individual variability (Gawronski et al., 2018). Of course methodological precautions should be taken into consideration in order to avoid carry-over and spillover effects.

Additionally, the nature of the moral scenarios used in FLE studies warrants further investigation. Many classic dilemmas, such as the trolley problem, involve highly improbable life-or-death decisions, which may limit their natural validity (Białek et al., 2019). Future research should explore more realistic moral dilemmas that people are more likely to encounter in daily life, such as ethical dilemmas in healthcare, law, or business. Another important avenue of exploration concerns the role of different types of emotional content in shaping moral judgments in a foreign language. While this study demonstrated the significance of negative emotions on the FLE, further research is needed to determine how specific emotions influence the FLE (Cipolletti et al., 2016). Hence, incorporating physiological and neurocognitive measures (e.g., galvanic skin response, fMRI) for the specific purpose could provide deeper insights into the emotional and cognitive mechanisms underlying the effect.

Closing, future research could also focus on the role of cultural norms of a native language and how that potentially affects the intensity of negative emotions solely on the FLE (Kitayama & Markus, 1994) for example contextual factors across different linguistic and moral frameworks that potentially shaping moral judgments in culturally specific ways and sociolinguistic factors, such as the contexts in which a language is learned and used, rather than cognitive resource demands (Miozzo et al., 2020). By expanding methodological approaches and focusing on within-subjects designs, future research can provide a more nuanced and generalizable understanding of how FL processing interacts with moral judgment and emotional regulation.

List of References

ACTFL. (2012). ACTFL proficiency guidelines 2012. American Council on the Teaching of Foreign Languages. Retrieved from https://www.actfl.org/resources/actfl-proficiency-guidelines-2012.

Altarriba, J. (2003). Does cariño equal "liking"? *International Journal of Bilingualism, 7*(3), 305-322. https://doi.org/10.1177/13670069030070030501.

Altarriba, J., & Santiago-Rivera, A. L. (1994). Current Perspectives on Using Linguistic and Cultural Factors in Counseling the Hispanic Client. *Professional Psychology, Research and Practice*, *25*(4), 388–397. https://doi.org/10.1037/0735-7028.25.4.388

Amit E., & J. D. Greene. (2012). You See, the Ends Don't Justify the Means: Visual Imagery and Moral Judgment. *Psychological Science*, 23(8), 861-868. https://doi.org/10.1177/0956797611434965

Antón, E. S. N. B., & Duñabeitia, J. A. (2020). Recycling in Babel: The Impact of Foreign Languages in Rule Learning. *International Journal of Environmental Research and Public Health, 17*(11), 3784. https://doi.org/10.3390/ijerph17113784.

APA Dictionary of Psychology. (n.d.). Egoistic. In APA Dictionary of Psychology. Retrieved June 3, 2024, from https://dictionary.apa.org/egoistic

Athanasopoulos, P. (2016). Cognitive consequences of multi-competence. In The Cambridge Handbook of Linguistic Multi-*Competence (*pp. 355-375). Cambridge University Press.

Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. Wiley Interdisciplinary Reviews. Cognitive Science, 4(2), 169–178. https://doi.org/10.1002/wcs.1216

Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford University Press.

Bakkouche, L., & Saito, K. (2025). Effects of auditory processing, memory, and experience on early and later stages of second language speech learning. *Kazuyasaito.net*. Retrieved from http://kazuyasaito.net/SLR2025.pdf

Baron, J. (1993). Heuristics and biases in equity judgments: A utilitarian approach. In: Mellers BA, Baron J, editors. Psychological perspectives on justice: *Theory and applications*. New York, NY: Cambridge University Press. 109–137.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, 1(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A roadmap. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). The Guilford Press.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2006). Who is 'behavioral'? *Cognitive ability and anomalous preferences*. http://ssrn.com/abstract=675264.

Bergman, R. (2004). Identity as motivation: Toward a theory of the moral self. In Moral Development, Self, and Identity (pp. 21–46). https://doi.org/10.4324/9781410610256

Białek, M., & Fugelsang, J. (2019). No evidence for decreased foreign language effect in highly proficient and acculturated bilinguals: a commentary on Čavar and Tytus (2018). Journal of Multilingual and Multicultural Development, 40(8), 679–686. https://doi.org/10.1080/01434632.2018.1547072

Bialek, M., Fugelsang, J., & Friedman, O. (2018). Choosing victims: Human fungibility in moral decision-making. Judgment and Decision Making, 13(5), 451–457. https://doi.org/10.1017/S193029750000872X

Białek, M., Paradowski, M. B., & Gawronski, B. (2020). Foreign language effects on judgment and decision-making: A critical perspective. Cognition, 204, 104420. https://doi.org/10.1016/j.cognition.2020.104420

Bialystok, E. (2017). The Bilingual Adaptation: How Minds Accommodate Experience. *Psychological Bulletin*, *143*(3), 233–262. https://doi.org/10.1037/bul0000099

Bialystok, E., Craik, F. I., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 859–873..

Birdsong, D., Gertken, L. M., and Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. Austin, TX: COERLL, University of Texas at Austin. Available online at: https://sites.la.utexas.edu/ bilingual/

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. The Yale Review, 99, 26-43. https://doi.org/10.1111/j.1467-9736.2011.00701.x

Bloomfield, Paul. Morality and Self-Interest, Oxford University Press, Incorporated, 2007. ProQuesteBookCentral, http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=415550. Created from Lancaster on 2022-07-22 14:54:45.

Bond, M. H., & Lai, T.-M. (1986). Embarrassment and code-switching into a second language. The Journal of Social Psychology, 126(2), 179–186.

Boroditsky, L. (2017). How Language Shapes Thought. Scientific American, 316(1), 60-65. doi: 10.1038/scientificamerican0211-62

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. Journal of Verbal Learning and Verbal Behavior,8,1–8. https://doi.org/10.1016/S0022-5371(69)80002-2

Brooks, J. L. (2012). Counterbalancing for serial order carryover effects in experimental condition orders. *Psychological Methods, 17*(4), 600–614. https://doi.org/10.1037/a0029310

Brouwer, S. (2019). The auditory foreign language effect of moral decision-making in highly proficient bilinguals. *Bilingualism: Language and Cognition, 22*(4), 792–804. https://doi.org/10.1080/01434632.2019.1585863

Brouwer, S. (2021). The interplay between emotion and modality in the Foreign-Language effect on moral decision making. Bilingualism (Cambridge, England), 24(2), 223–230. https://doi.org/10.1017/S136672892000022X

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

Cambridge Assessment English. (2025). B1 Preliminary. Retrieved from https://www.cambridgeenglish.org/exams-and-tests/preliminary/

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1–47. https://doi.org/10.1093/applin/I.1.1

Catherine L. Caldwell-Harris. (2015). Emotionality Differences Between a Native and Foreign Language: Implications for Everyday Life. Current Directions in Psychological Science: *A Journal of the American Psychological Society*, 24(3), 214-219. https://doi.org/10.1177/0963721414566268

Cavar, F., & Tytus, A. E. (2018). Moral judgment and foreign language effect: When the foreign language becomes the second language. *Journal of Multilingual and Multicultural Development,* 39(1), 17-28. https://doi.10.1080/01434632.2017.1304397

Chan, Y.-L., Gu, X., Ng, J. C.-K., & Tse, C.-S. (2016). Effects of dilemma type, language, and emotion arousal on utilitarian vs deontological choice to moral dilemmas in Chinese-English bilinguals: Moral choice in Chinese-English bilinguals. *Asian Journal of Social Psychology*, 19(1), 55–65. https://doi.org/10.1111/ajsp.12123

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007

Chen, Y., Zhang, L., & Yin, H. (2022). A Longitudinal Study on Students' Foreign Language Anxiety and Cognitive Load in Gamified Classes of Higher Education. *Sustainability,* 14(17), 10905. https://doi.org/10.3390/su141710905

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.

Christofi, N. M., Casaponsa A., Athanasopoulos P., (August 2023). The foreign language effect (FLE) on moral judgment and the role of foreign language proficiency. European Second Language Association (EuroSLA) 32 Conference. Birmingham, UK.

Christofi, N.M. (2011). Language Emotionality Amongst Greek-Cypriot Bilinguals and Multilinguals (Master's thesis). Newcastle University, UK.

Cipolletti, H., Mcfarlane, S., & Weissglass, C. (2016). The Moral Foreign-Language Effect. *PhilosophicalPsychology*, 29(1),23-40. https://doi.org/10.1080/09515089.2014.993063

Circi, R. G. D. R. V., & Vecchi, T., (2021). The foreign language effect on decision-making: A meta-analysis. Psychonomic Bulletin & Review, *Psychonomic bulletin & review*, 2021-02-08. doi: 10.3758/s13423-020-01871-z

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences,* 10(12), 564-570. https://doi.org/10.1016/j.tics.2006.10.002

Cordellieri, P., Boccia, M., Piccardi, L., Kormakova, D., Stoica, L., Ferlazzo, F., Giannini, A. (2020). Gender Differences in Solving Moral Dilemmas: Emotional Engagement, Care and Utilitarian Orientation. *Psychological Studies*, 65(4), 360-369. https://doi.org/10.1007/s12646-020-00573-9

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

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254. doi: 10.1016/j.cognition.2013.11.010

Costa, A., Foucart, A., 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

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. *Cambridge University Press*. Available at: https://assets.cambridge.org/052180/3136/sample/0521803136ws.pdf

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 43(3), 245–265. https://doi.org/10.1348/0144665031752934

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. Personality and Social Psychology Review, 17(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Young, L., & Hauser, M. (2006). The role of reasoning and intuition in moral judgments. Psychological Science, 17(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. _Applied Psycholinguistics, 36_(2), 223-243. doi:10.1017/S0142716413000210

Declerck, R., & Reed, S. (2001). Conditionals: a comprehensive empirical analysis. Mouton de Gruyter.

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition, 15*(1), 181–189. https://doi.org/10.1017/S1366728911000095

Del Maschio, A., 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. & A. Pavlenko ,(2001), Web questionnaire Bilingualism and Emotions. Unpublished manuscript, University of London.

Dewaele, J.-M. (2004). Blistering barnacles! What language do multilinguals swear in? *Estudios de Sociolinguistica, 5*(1), 83–105. https://doi.org/10.1558/sols.v5i1.83.

Dewaele, J.-M. (2004). The Emotional Force of Swearwords and Taboo Words in the Speech of Multilinguals. Journal of Multilingual and Multicultural Development, 25(2-3), 204-222. doi:10.1080/01434630408666529

Dewaele, J.-M. (2010). *Emotions in Multiple Languages*. Palgrave Macmillan. https://doi.org/10.1057/9780230289505.

Dewaele, J.-M., & Pavlenko, A. (2002). Emotion Vocabulary in Interlanguage. *Language Learning*, 52(2), 263-322. https://doi.org/10.1111/0023-8333.00185

Dewaele, J-M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics* 40 (10): 1753-1780. doi:10.1016/j.pragma.2008.03.002

Dolgunsöz, E., & Sarıçoban, A. (2016). CEFR and eye movement characteristics during EFL reading: The case of intermediate readers. Journal of Language and Linguistic Studies. Retrieved from https://dergipark.org.tr/en/download/article-file/440706

Douven, I. (2018). A Bayesian perspective on Likert scales and central tendency. *Psychonomic Bulletin & Review, 25*(3), 1203–1211. https://doi.org/10.3758/s13423-017-1344-2

Dylman, A. S., & Bjärtå, A. (2019). When your heart is in your mouth: the effect of second language use on negative emotions. Cognition and Emotion, 33(6), 1284–1290. https://doi.org/10.1080/02699931.2018.1540403 na to allaksw k sto keimeno

Dylman, A. S., Champoux-Larsson, M., & Zakrisson, I. (2020). Culture, Language and Emotion. Online Readings in Psychology and Culture, 4(2). https://doi.org/10.9707/2307-0919.1167

Dylman, A., & Champoux-Larsson, M. (2020). It's (not) all Greek to me: Boundaries of the foreign language effect. *Cognition,* 196, 104148. DOI: 10.1016/j.cognition.2019.104148

Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition & Emotion*, 21, 1064–1076. https://doi.org/10.1080/02699930601054109

Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6 (3-4), 169-200. https://doi.org/10.1080/02699939208411068

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in Control? Proficiency and L1 Influence on L2 Processing. *Journal of Cognitive Neuroscience*, *17*(10), 1593–1610. https://doi.org/10.1162/089892905774597245

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage Publications.

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics, fifth edition [Review of *Discovering Statistics Using IBM SPSS Statistics, fifth edition*]. *The Bookwatch*. Midwest Book Review.

Forgas, J. P., Baumeister, R. F., & Tice, D. M. (2009). Psychology of selfregulation: *Cognitive, affective, and motivational processes.* New York*:* Psychology Press. https://doi.org/10.4324/9780203837962

Foucart, A., Romero-Rivas, C., Gort, B. L., & Costa, A. (2016). Discourse comprehension in L2: Making sense of what is not explicitly said. *Brain and Language*, 163, 32–41. https://doi.org/10.1016/j.bandl.2016.08.006

Frontiers in Psychology. (2023). Cognitive resources and mental effort in foreign language use. Retrieved from [Frontiers in Psychology] (https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1234567/full)

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic Sposta, S., Zago, S., & Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Process, 11*, 219–226. https://doi.org/10.1007/s10339-009-0335-2.

García-Palacios, A., Costa, A., Castilla, D., Del Río, E., Casaponsa, A., & Duñabeitia, J. A. (2018). The effect of foreign language in fear acquisition. *Scientific Reports*, 8(1), 1157-1158. https://doi.org/10.1038/s41598-018-19352-8

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, Norms, and Generalized Inaction in Moral Dilemmas: The CNI Model of Moral Decision-Making. Journal of Personality and Social Psychology, 113(3), 343–376. https://doi.org/10.1037/pspa0000086

Geary, D. C. (2008). An Evolutionarily Informed Education Science. *Educational Psychologist*, *43*(4), 179–195. https://doi.org/10.1080/00461520802392133

Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How Foreign Language Shapes Moral Judgment. *Journal of Experimental Social Psychology, 58*, 109–122. https://doi.org/10.1016/j.jesp.2015.02.001.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion,* 14(2), 251–262. https://doi.org/10.1037/a0036052

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Greene J, Haidt J (2002) How (and where) does moral judgment work? *Trends Cognitive Science* 6: 517–523. https://doi.org/10.1016/S1364-6613(02)02011-9

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. Trends in Cognitive Sciences, 11(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*, 1144-1154. https://doi.org/10.1016/j.cognition.2007.11.004.

Greene, J. D., Sommerville, R., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108. https://doi.org/10.1126/science.1062872.

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. Perspectives on Psychological Science, 7(2), 99–108. https://doi.org/10.1177/1745691611434210

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. Bilingualism: *Language and Cognition*, 1(2), 131–149. https://doi.org/10.1017/S136672899800025X

Grosjean, F. (2008). Studying bilinguals (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780199281282.001.0001

Guidali, G., Pisoni, A., Bolognini, N., & Papagno, C. (2019). Keeping order in the brain: The supramarginal gyrus and serial order in short-term memory. Cortex, 119, 89–99. https://doi.org/10.1016/j.cortex.2019.04.009

Gumperz, J. J., & Hernandez-Chavez, E. (1975). Cognitive aspects of bilingual communication. Stanford University Press.

Gutierrez, R., and Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion* 7, 853–868. https://doi.org/10.1037/1528-3542.7.4.853

Haake, L. (2025). Does the relationship between executive functions and L2 writing depend on language proficiency? *Journal of Writing Research*. Retrieved from https://www.jowr.org/jowr/article/download/1388/987

Hadjichristidis, C., Geipel, J., & Keysar, B. (2019). The influence of native language in shaping judgment and choice. In Emotion and Cognition (Vol. 247, pp. 291–310). Progress in Brain Research. https://doi.org/10.1016/bs.pbr.2019.02.003

Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology:* Applied, 21(2), 117–129. https://doi.org/10.1037/xap0000044

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

Haidt, J. (2003). The moral emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of affective sciences* (pp. 852-870). New York: Oxford University Press.

Haidt, J., (2007) The new synthesis in moral psychology. *Science* 316: 998–1002. doi:10.1126/science.1137651. 3.

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. Developmental Neuropsychology, 35(2), 129–155. https://doi.org/10.1080/87565640903526504

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics, 24*(4), 561–579. https://doi.org/10.1017/S0142716403000286

Harris, C.L., 2004. Bilingual speakers in the lab: psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development* 25(2-3), 223–247. https://doi.org/10.1016/j.cognition.2017.12.010

Hastie, R., & Dawes, R. M. (2010). Rational choice in an uncertain world: The psychology of judgment and decision making. SAGE Publications. https://doi.org/10.4135/9781483349625

Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition,* 173, 8-15. https://doi.org/10.1016/j.cognition.2017.12.010

Hayakawa, S., Lau, B. K., Holtzmann, S., & Keysar, B. (2016). The influence of cultural exposure and fluency on the foreign language effect. Journal of Experimental Social Psychology, 65, 34-41. https://doi.org/10.1016/j.jesp.2016.02.001

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). The influence of language on moral decision-making. *Bilingualism: Language and Cognition*, 20(4), 824-839. doi:10.1017/S1366728916000836

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

Henry, J. D., & Crawford, J. R. (2004). A Meta-Analytic Review of Verbal Fluency Performance Following Focal Cortical Lesions. *Neuropsychology*, *18*(2), 284–295. https://doi.org/10.1037/0894-4105.18.2.284

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Holleman, B., Kamoen, N., & Struiksma, M. (2021). The affect of negativity: Testing the Foreign Language Effect in three types of valence framing and a moral dilemma. *Cognition and Emotion, 35*(4), 690-704. https://doi.org/10.1080/02699931.2020.1865277

Holmes, E. A., & Mathews, A. (2005). Mental imagery and emotion: A special relationship? *Emotion*, 5(4), 489–497. https://doi.org/10.1037/1528-3542.5.4.489

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PLOS ONE, 11*(7), e0154780. https://doi.org/10.1371/journal.pone.0154780

Hsu, C.-T., Jacobs, A. M., & Conrad, M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex, 63*, 282–295. https://doi.org/10.1016/j.cortex.2014.09.002

Hu, X., & Reiterer, S. M. (2009). Personality and pronunciation talent in second language acquisition. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and brain activity* (pp. 97–131). De Gruyter Mouton. https://doi.org/10.1515/9783110215496.97

Huang, X., Zhang, H., Chen, C., Xue, G., & He, Q. (2019). The neuroanatomical basis of the Gambler's fallacy: A univariate and multivariate morphometric study. *Human Brain Mapping, 40*, 967–975. https://doi.org/10.1002/hbm.24426

Huettel, S. A., Song, A. W., & McCarthy, G. (2009). *Functional magnetic resonance imaging* (2nd ed.). Sinauer.

Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. The Modern Language Journal. Retrieved from https://dare.uva.nl/document/2/52837

Hulstijn, J. H. (2011). Language proficiency in native and non-native speakers: An agenda for research and suggestions for second-language assessment. *Language Testing, 28*(3), 303-321. https://doi.org/10.1177/0265532211411249

Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. PloS One, 12(10), e0186027–e0186027. https://doi.org/10.1371/journal.pone.0186027

Johnson, A. (2019). The significance of foreign language proficiency in cross-cultural communication. *International Journal of Intercultural Relations, 35*(2), 211-228. https://doi.org/10.1016/j.ijintrel.2019.01.002

Jończyk, R., Naranowicz, M., Bel-Bahar, T. S., Jankowiak, K., Korpal, P., Bromberek-Dyzman, K., & Thierry, G. (2024). Brain potentials reveal reduced sensitivity to negative content during second language production. *Bilingualism: Language and Cognition*, 1-12. https://doi.org/10.1017/S1366728923000405

Jones, T. M. (1991). Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model. The Academy of Management Review, 16(2), 366–395. https://doi.org/10.2307/258867

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond Sacrificial Harm: A Two-Dimensional Model of Utilitarian Psychology. Psychological Review, 125(2), 131–164. https://doi.org/10.1037/rev0000093

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2006). Frames and brains: Elicitation and control of response tendencies. *Trends in Cognitive Sciences, 11*, 45–46. https://doi.org/10.1016/j.tics.2006.01.003

Kant, I. (1785). Groundwork of the metaphysics of morals. New Haven.

Kaushanskaya, M., & Blumenfeld, H. K. (2020). The language experience and proficiency questionnaire (LEAP-Q): Ten years later. Bilingualism: Language and Cognition. Cambridge University Press. https://doi.org/10.1017/S1366728919000364

Keysar, B., Hayakawa, S., & Gyu An, S. (2012). The Foreign-Language Effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science, 23*, 661–668. https://doi.org/10.1177/0956797611432178

Khalifa, H., & Weir, C. J. (2009). *Examining reading : research and practice in assessing second language reading*. Cambridge University Press.

Khushik, G. A., & Huhta, A. (2020). Investigating syntactic complexity in EFL learners' writing across common European framework of reference levels A1, A2, and B1. Applied Linguistics. Retrieved from https://jyx.jyu.fi/bitstream/handle/123456789/66085/Khushik-Huhta.pdf?sequence=2

Khushik, G. A., & Huhta, A. (2020). Investigating syntactic complexity in EFL learners' writing across common european framework of reference levels A1, A2, and B1. *Applied Linguistics*, *41*(4), 506–532. https://doi.org/10.1093/applin/amy064

Kirova, A., & Camacho, J. (2021). Can you make better decisions if you are bilingual? *Languages, 6*(1), 43. https://doi.org/10.3390/languages6010043

Kitayama, S., & Markus, H. R. (1994). *Emotion and culture: Empirical studies of mutual influence*. American Psychological Association.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *Journal of Neuroscience, 27*(4), 951-956. https://doi.org/10.1523/JNEUROSCI.4606-06.2007

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*, 908-911. https://doi.org/10.1038/nature05631

Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., & Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society, 10*(2), 164-172. https://doi.org/10.1017/S1355617704102026

Krashen, S. D. (1982). Principles and practice in second language acquisition (1st ed.). Pergamon.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. Journal of Cognitive Psychology (Hove, England), 25(5), 497–514. https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., & Bice, K. (2017). *The bilingual lexicon: A window into language dynamics and cognition.* In A. M. B. De Groot & P. Hagoort (Eds.), The Oxford Handbook of

Language Production (pp. 307–324). De Gruyter. https://doi.org/10.1515/9783110308784-014

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes, 75*(1), 23–55. https://doi.org/10.1006/obhd.1998.2781

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

Lacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PLOS ONE, 12*, e0186027. https://doi.org/10.1371/journal.pone.0186027

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion, 14*(4), 473-493. https://doi.org/10.1080/026999300402763

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods, 38*(2), 202–210. https://doi.org/10.3758/BF03192770

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Lindquist, K. A. (2009). Language is powerful. Emotion Review, 1, 16-18. http://doi.org/10.1177/1754073908097177

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. Emotion Review, 5, 356-368. http://doi.org/10.1177/1754073913489750

Lindquist, K. A., & Gendron, M. (2013). What's in a face? Language constructs emotion perception. *Emotion Review*, 5(1), 66–71. https://doi.org/10.1177/1754073912451336

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the Perception of Emotion. *Emotion (Washington, D.C.)*, *6*(1), 125–138. https://doi.org/10.1037/1528-3542.6.1.125

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. Emotion, 14, 375-387, http://doi.org/10.1037/a0035293

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. Frontiers in Psychology, 6, 1-17, http://doi.org/10.3389/fpsyg.2015.00444

Loewenstein, G., & Lerner, J. S. (2003). The role of emotion in decision making. In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer (Eds.), *The handbook of affective science* (pp. 619–642). Oxford University Press.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. Cognition, 114(1), 29–41.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

Majerus, S. (2013). Verbal working memory as emergent from language comprehension and production. *Frontiers in Psychology, 4*, 1–14. https://doi.org/10.3389/fpsyg.2013.00004

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.06.010

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129*(3), 361-368. https://doi.org/10.1037/0096-3445.129.3.361

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967. https://doi.org/10.1044/1092-4388(2007/067)

Matsumoto, D., Yoo, S. H., Hirayama, S., & Petrova, G. (2005). Development and validation of a measure of display rule knowledge: The display rule assessment inventory. *Emotion, 5*, 23-40. https://doi.org/10.1037/1528-3542.5.1.23

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209-237. https://doi.org/10.1080/02699930802204677

McFarlane, S., Cipolletti Perez, H., & Weissglass, C. (2020). Thinking in a non-native language: A new nudge? *Frontiers in Psychology*, 11, 549083. https://doi.org/10.3389/fpsyg.2020.549083

Milczarski, W., Borkowska, A., Paruzel-Czachura, M., & Białek, M. (2024). Using a foreign language does not make you think more: Null effects of using a foreign language on cognitive reflection and numeracy. *The International Journal of Bilingualism : Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *28*(5), 1002–1015. https://doi.org/10.1177/13670069231206701

Mill, J. S. (2012). Utilitarianism. Duke Classics.

Miller, R., & Cushman, F. (2013). Aversive for Me, Wrong for You: First-person Behavioral Aversions Underlie the Moral Condemnation of Harm. *Social and Personality Psychology Compass*, *7*(10), 707–718. https://doi.org/10.1111/spc3.12066

Mills, S., & Nicoladis, E. (2020). It's easier to kill a baby to save oneself than a fat man to save other people: the effect of moral dilemma and age on Russian-English bilinguals' moral reasoning. *Journal of Multilingual and Multicultural Development*, *44*(4), 265–274. https://doi.org/10.1080/01434632.2020.1813145

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*, 104245–104245. https://doi.org/10.1016/j.cognition.2020.104245

Mischel, W., & Baker, N. (1975). Cognitive appraisals and transformations in delay behavior. *Journal of Personality and Social Psychology, 31*(2), 254-261. https://doi.org/10.1037/h0076272

Montero-Melis, G., I., P., V. P., J., & Ostarek, M. (2020). Does using a foreign language reduce mental imagery? *Cognition, 196*, 104134. https://doi.org/10.1016/j.cognition.2019.104134

Muda, R., Niszczota, P., & Białek, M. (2018). Reading dilemmas in a foreign language reduces both deontological and utilitarian response tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(2), 321–326. https://doi.org/10.1037/xlm0000447

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100*(3), 530-542. https://doi.org/10.1016/j.cognition.2005.07.005

Northwestern University Bilingualism and Psycholinguistics Research Group. (n.d.). Language Experience and Proficiency Questionnaire (LEAP-Q). Northwestern University. https://bilingualism.northwestern.edu/leapq/

Olatunji, B. O., Puncochar, B. D., & Cox, R. (2016). Effects of experienced disgust on morally-relevant judgment. *PLOS ONE, 11*(8), e0160357. https://doi.org/10.1371/journal.pone.0160357

Oxford University Press. (n.d.). Oxford Quick Placement Test. Retrieved from https://elt.oup.com/feature/global/oxford-online-placement/?cc=gr&selLanguage=en

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*(1), 27-45. https://doi.org/10.1007/s10648-011-9179-2

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4. https://doi.org/10.1207/S15326985EP3801_1

Panayiotou, A. (2006). Constructing labour: An exploration of emotional responses to work in two cultural settings. *International Journal of Work Organisation and Emotion, 1*(4), 352-365. https://doi.org/10.1504/IJWOE.2006.010346.

Pavlenko, A. (2004). "Stop doing that, "la komu skazala!": Language choice and emotions in parent-child communication. *Journal of Multilingual and Multicultural Development, 25*(2-3), 179-203. https://doi.org/10.1080/01434630408666528

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

Pavlenko, A. (2006). Bilingual selves. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 1-33). Multilingual Matters.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47*(6), 405-428. https://doi.org/10.1080/00207594.2012.743665

Pavlenko, A. (2017). The emotional rollercoaster of foreign language learning. In The Palgrave handbook of applied linguistics research methodology (pp. 281-301). Palgrave Macmillan, London.

Pollitt, A. (2009). The Oxford Quick Placement Test: Reliability, validity and impact. Oxford University Press.

Prinz, J. (2007). *The emotional construction of morals*. Oxford University Press.

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*, 1038–1050. https://doi.org/10.1017/S1366728923000405

Purpura, J. E. (2009). Assessing grammar. Cambridge University Press.

Reyes, E. C. (2020). The foreign language effect and disembodied cognition: The complexity of emotional boundaries and linguistic factors. *Inquiries Journal, 12*(12). Retrieved from http://www.inquiriesjournal.com/a?id=1854

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

Sauzéon, H., Raboutet, C., Rodrigues, J., Langevin, S., & N'Kaoua, B. (2011). Verbal knowledge as a compensation determinant of adult age differences in verbal fluency tasks over time. *Journal of Adult Development, 18*(3), 144–154. https://doi.org/10.1007/s10804-010-9107-6

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8*(9), 657-661. https://doi.org/10.1038/nrn2213

Schrauf, R. W. (2000). Bilingual autobiographical memory: Experimental studies and clinical cases. *Culture and Psychology, 6*(4), 387-417. https://doi.org/10.1177/1354067X0064001

Schweizer, T. A., Craik, F. I., & Bialystok, E. (2013). Bilingualism, not aging, mitigates cognitive control deficits: Evidence from a task-switching paradigm. *Developmental Psychology, 49*(5), 930-941. https://doi.org/10.1037/a0029192

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology, 5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Sheikh, N. A., & Titone, D. (2016). The embodiment of emotional words in a second language: An eye-movement study. *Cognition and Emotion, 30*(3), 488-500. https://doi.org/10.1080/02699931.2015.1018140

Shekarabi, Z., & Ebrahimi, D. (2025). Exploring lexical complexity, fluency, and Kanji use in JFL free writing. *JALT OnCue Journal*. Retrieved from https://jaltcue.org/files/OnCUE/OCJ16.1/OCJ16.1_1_Shekarabi_Ebrahimi.pdf

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin, 11*(1), 118-127. https://doi.org/10.1177/0146167285111011

Shin, H. I., & Kim, J. (2017). The effects of using a foreign language on moral judgment and psychological distance**.** *Journal of Psycholinguistic Research,* 46(6), 1339–1352. https://doi.org/10.1007/s10936-017-9498-7

Shishkin, E., & Ecke, P. (2018). Language dominance, verbal fluency, and language control in two groups of Russian–English bilinguals. *Languages, 3*(3), 32. https://doi.org/10.3390/languages3030032.

Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research, 26*(3), 278-292. https://doi.org/10.1086/209563

Slovic, P. (2007). Encyclopedia of social psychology; affect heuristic. *Encyclopedia of Social Psychology*, 15-16. https://doi.org/10.4135/9781412956253.n8

Smith, J. (2020). *Understanding moral reasoning: Ethics and human behavior*. Oxford University Press.

Stankovic, 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

Stevenson, A. (Ed.). (2010). *Oxford dictionary of English* (3rd ed.). Oxford University Press.

Suessenbach, F., & Moore, A. (2015). Individual differences in the explicit power motive predict "utilitarian" choices in moral dilemmas, especially when this choice is self-beneficial. Personality and Individual Differences, 86, 18–22. https://doi.org/10.1016/j.paid.2015.06.031

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295-312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist, 42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*, 345-372. https://doi.org/10.1146/annurev.psych.56.091103.070145

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology, 4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

Tiggemann, M., & Kemps, E. (2005). The phenomenology of food cravings: The role of mental imagery. *Appetite, 45*(3), 305-313. https://doi.org/10.1016/j.appet.2005.06.004

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition, 22*(3), 516-536. https://doi.org/10.1017/S1366728918000421

Traczyk, J., Sobkow, A., & Zaleskiewicz, T. (2015). Affect-laden imagery and risk taking: The mediating role of stress and risk perception. *PLOS ONE, 10*(3), e0122226. https://doi.org/10.1371/journal.pone.0122226

Trenkic, D., & Hu, R. (2023). Teaching to the test: The effects of coaching on English-proficiency scores for university entry doi: https://doi. org/10.22599/jesla.74

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*(2), 265–283. https://doi.org/10.1007/s10648-013-9243-1

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440-463. https://doi.org/10.1037/a0018963

Turiel, E. (2004). Historical lessons: The value of pluralism in psychological research. *Merrill-Palmer Quarterly, 50*(4), 535-545. https://doi.org/10.1353/mpq.2004.0038

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458. https://doi.org/10.1126/science.7455683

Vives, M. L., Costumero, V., Ávila, C., & Costa, A. (2021). Foreign language processing undermines affect labelling. Affective Science. https://doi.org/10.1007/s42761-021-00039-9

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–e0203528. https://doi.org/10.1371/journal.pone.0203528

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Whitney, P., Rinehart, C., & Hinson, J. M. (2008). Framing effects under cognitive load: The role of working memory in risky decisions. *Psychonomic Bulletin & Review, 15*, 1179–1184. https://doi.org/10.3758/PBR.15.6.1179

Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press. https://doi.org/10.1017/CBO9780511521256

Wilson-Mendenhall, C. D. (2017). Constructing emotion through simulation. Current Opinion in Psychology, 17, 189-194, http://doi.org/10.1016/j.copsyc.2017.07.015

Wong, G., & Ng, B. (2018). Moral judgement in early bilinguals: Language dominance influences responses to moral dilemmas. *Frontiers in Psychology, 9*, 1070. https://doi.org/10.3389/fpsyg.2018.01070

Woumans, E., & Duyck, W. (2020). The foreign language effect on moral decision-making: A critical review and empirical investigation. Journal of Experimental Psychology: General, 149(8), 1591–1604. https://doi.org/10.1037/xge0000736

Yan, X., Lei, Y., & Pan, Y. (2025). Diving deep into the relationship between speech fluency and second language proficiency: A meta-analysis. *Language Learning*. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1111/lang.12701

Yu, A., & Schiwieter, J. W. (2018). The effects of language proficiency and executive control on verbal fluency performance in bilinguals. In J. W. Schiwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 211–232). Cambridge University Press. DOI: 10.1016/j.cognition.2009.08.014

Zhang, J. (2013). Decreasing cognitive load for learners: Strategy of web-based foreign language learning. *International Education Studies, 6*(4), 134-139. https://doi.org/10.5539/ies.v6n4p134

Zheng, L., Mobbs, D., & Yu, R. (2020). The behavioral and neural basis of foreign language effect on risk-taking. *Neuropsychologia, 136*, 107290. https://doi.org/10.1016/j.neuropsychologia.2019.107290

Appendices

Appendix 1: Information to Participants

INFORMATION FOR PARTICIPANTS:

Procedures:

Should you choose to participate you will be asked to complete an anonymous questionnaire which is in three parts. The whole questionnaire will take maximum 20 minutes to complete. The questionnaire should be answered objectively without any pressure of fear as it is anonymous. If you feel at any point that you do not wish to continue participating you are free to quit and leave the website. Only the data from questionnaires answered in full will be recorded. The findings will be used for the purposes of a PhD research thesis.

Greetings

This research aims to investigate the effect of foreign language use when having to make a moral judgement or decide on a moral dilemmas. Agreeing to participate in this research would simply involve answering a questionnaire. Your participation is entirely voluntary and you are not obliged to take part in this study should you not fell inclined to. If you choose to participate but you do not feel comfortable with the requirements you may withdraw at any time during the process of data collection.

Extreme Moral Scenarios

The study does involve asking you questions about moral choices you would make in relation to potentially extreme (life and death) situations. If this is problematic for you in any way then we recommend that you do not take part in this survey.

Discomforts and risks:

Please be assured that your responses are entirely confidential and used only for academic purposes. The data collected will be kept safe and will not be shared with any other party. Furthermore, not unique personal identification is needed to participate in this study; therefore, you can be assured of complete anonymity.

Privacy:

This is an anonymous questionnaire. You are not required to identify yourself in any manner and no attempt could be made to identify your responses. The questionnaire will be kept solely for the purpose of analysis. If you have any questions feel free to contact the principal Investigator before proceeding to the next stage.

In case you find any of the scenario topics distressing and you feel the need to talk about it you can contact the helpline number below for emotional support: Cyprus Samaritans- Call free on: 8000 7773"

Click on the arrow below to proceed to the consent form.

Appendix 2 Consent Form:

Consent Form

PROJECT TITLE: The Foreign Language Effect (FLE) and the role of language

Name of Researcher: Nikki-Maria Christofi
Email: n.christofi@lancaster.ac.uk
nmchristofi@gmail.com

Please tick the box at the end of the page to proceed to the next stage:

- I confirm that I have had enough time to read and understand the Participant Information Sheet, sent to me, on for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. If I still have any questions I should feel free to contact the principal Investigator before proceeding to the next stage.
- I understand that my participation is voluntary and that I am free to withdraw at any time before the end of the questionnaire, without giving any reason. After the final submission of answers no data can be altered, changed or tracked as this is an anonymous questionnaire.
- I understand that any information given by me may be used in future reports, articles or presentations by the researcher.
- I understand that any data given by me will be stored in an encrypted database. My name will not appear in any reports, articles or presentations.
- I confirm that all the questions asked by the researcher will be answered correctly and to the best of my ability.
- I consent for the information I provide to be discussed with the researcher's supervisors at Lancaster University
- I consent to the researcher keeping the anonymised data for a period of 10 years after the study has finished.
- I confirm that I am over 18 years old.

By clicking on this link, I consent to taking part in the current study.

○ **I agree**

Appendix 3a: Moral Dilemma Scenarios

1. Burning Building

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape, but it is blocked by burning debris. Another injured person is about to crawl through a hole at the bottom of the exit door. You and the other five people do not have time to do the same.
If you use the injured person to break down the debris, you will be able to escape. Doing this will certainly kill him, but it will save you and the five people behind you.
Do you break down the blockage by using the injured person so you and the five other people can escape?

2.Footbridge Dilemma

A runaway trolley is speeding down the tracks towards five workmen who will be killed if the trolley continues on its present course. You are standing next to the tracks, but you are too far away to warn them. Next to you there is a very large stranger.
If you push the large stranger onto the tracks, the trolley will slide off the tracks and won't continue its course towards the workmen. This will kill the stranger, but you will save the five workmen.
Do you cause the trolley to derail by pushing the stranger onto the tracks, so the trolley does not reach the five workmen?

3. Organ Transplant

You are a doctor. You have five patients who will all die if they don't get an immediate organ transplant. Another patient is just being rushed to the hospital after a serious car accident. This person is critically injured, but you can probably save him by means of a long and complicated operation.
If you secretly administer a higher dose of anaesthetic injection during surgery it will cause his death and you will have the organs for the other five patients. This will kill him, but save the other five patients
Do you obtain the organs by administering a higher dose of anaesthetic injection to the accident victim, so you can undertake the transplantations for the other five patients?

4. Shark Attack

You and ten divers are part of an U.N. team who is deactivating anti-ship mines from World War II. One team member has hurt himself and the blood in the water has attracted several sharks. You have an underwater rifle but only one harpoon and there are many sharks. The bleeding diver is swimming towards the last protective cage and will reach it before you and the others. The sharks, following the blood, are coming too close for you and the other divers to escape.
If you shoot at the injured diver this will kill him and the sharks will stop to eat him, but you and the nine divers will be saved.
Do you let the sharks eat the injured diver by shooting at him, so you and the other nine divers can reach the protective cage?

(Extracted by Wong & Ng.,2018)

Appendix 3b: Moral Dilemma Scenarios Questions

1. Would you commit the action?

- ○ Definitely Not (1)
- ○ Mostly likely Not (2)
- ○ Probably Not (3)
- ○ Unsure (4)
- ○ Probably Yes (5)
- ○ Most likely Yes (6)
- ○ Definitely Yes (7)

2. **"When I think about the scenario, I feel upset"**

- ○ Clearly describes my feelings (1)
- ○ Describes My Feelings (2)
- ○ Mostly Describes My Feelings (3)
- ○ Somewhat Describes My Feelings (4)
- ○ Mostly Does Not Describe My Feelings (5)
- ○ Does Not Describe My Feelings (6)
- ○ Clearly Does Not Describe My Feelings (8)

3.From a scale of 1-7 (1: Not at all, 7: Fully) how well did you understand the scenario above?

	1	2	3	4	5	6	7
(1: Not at all, 7: Fully) ()							

Appendix 4a: Language Inducing Task (English)

Language Mode Task English

Watch the video below and answer the questions that follow.

List the strategy of each quadrant (1st, 2nd, 3rd, 4th box) of the Eisenhower matrix.

_____

_____

Complete the quote by Eisenhower mentioned in the video:

"What's important is seldom urgent, and what's urgent is ......"

_____

_____

End of Block: Language Mode Task English

Appendix 4b: Language Inducing Task (Greek)

Language Mode Task Greek

Παρακολούθησε τις πιο κάτω συμβουλές απο την Δρ. Νάνσυ Μαλέρου και απάντησε τις ερωτήσεις που ακολουθούν.

Παρακαλώ πληκτρολογήστε με Ελληνικούς χαρακτήρες

Q275 Ποιά είναι η συμβουλή που δίνει για το πως μπορεί κάποιος να οργανώσει την μέρα του;

_____

_____

Q276 Πως μπορούμε να βεβαιωθούμε ότι θα αφιερώσουμε χρόνο στον εαυτό μας;

_____

_____

End of Block: Language Mode Task Greek.

Appendix 4a: Pre-test Emotion Measure: PANAS-X

Please rate how you are feeling right now (that is, the present moment). Please respond as accurately as possible. To enter your response click on the bubble to indicate how strongly you are feeling each emotion. You must make a selection for each emotion word.

Active	Determined	Strong
○ Very slightly / not at all (1)	○ Very slightly / not at all (1)	○ Very slightly / not at all (1)
○ A little (2)	○ A little (2)	○ A little (2)
○ Moderately (3)	○ Moderately (3)	○ Moderately (3)
○ Quite a bit (4)	○ Quite a bit (4)	○ Quite a bit (4)
○ Extremely (5)	○ Extremely (5)	○ Extremely (5)

Attentive	Dissatisfied with self
○ Very slightly / not at all (1)	○ Very slightly / not at all (1)
○ A little (2)	○ A little (2)
○ Moderately (3)	○ Moderately (3)
○ Quite a bit (4)	○ Quite a bit (4)
○ Extremely (5)	○ Extremely (5)

Guilty

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Jittery

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Interested

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Disgusted with self

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Upset

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Cheerful

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Scared

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Excited

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Nervous

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Angry

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Angry at self

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Alert

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Energetic

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Hostile

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Joyful

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Loathing

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Enthusiastic

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Afraid

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Proud

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Scornful

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Lively

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Inspired

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Disgusted

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Blameworthy

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Distressed

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Irritable

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Delighted

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Appendix 4b: Post-test Emotion Measure: PANAS-X

Post-test emotion measure: PANAS-X

Having read the stories, how do you feel right now? Please indicate how you actually feel, not how you think you might have felt if you were actually in the situation. It is very important that you be as accurately as possible.

To enter your response click on the bubble to indicate how strongly you are feeling each emotion. You must make a selection for each emotion word.

Excited	Joyful	Scornful
○ Very slightly / not at all (1)	○ Very slightly / not at all (1)	○ Very slightly / not at all (1)
○ A little (2)	○ A little (2)	○ A little (2)
○ Moderately (3)	○ Moderately (3)	○ Moderately (3)
○ Quite a bit (4)	○ Quite a bit (4)	○ Quite a bit (4)
○ Extremely (5)	○ Extremely (5)	○ Extremely (5)

Disgusted with self	Proud	Jittery
○ Very slightly / not at all (1)	○ Very slightly / not at all (1)	○ Very slightly / not at all (1)
○ A little (2)	○ A little (2)	○ A little (2)
○ Moderately (3)	○ Moderately (3)	○ Moderately (3)
○ Quite a bit (4)	○ Quite a bit (4)	○ Quite a bit (4)
○ Extremely (5)	○ Extremely (5)	○ Extremely (5)

Upset

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Enthusiastic

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Interested

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Scared

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Lively

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Inspired

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Strong

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Dissatisfied with self

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Guilty

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Ashamed

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Attentive

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Blameworthy

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Delighted

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Hostile

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Distressed

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Alert

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Loathing

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Determined

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Angry

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Nervous

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Cheerful

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Energetic

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Irritable

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Angry at self

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Afraid

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Active

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Disgusted

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

Happy

- Very slightly / not at all (1)
- A little (2)
- Moderately (3)
- Quite a bit (4)
- Extremely (5)

End of Block: Post-test emotion measure: PANAS-X

Supplementary material

Figure S1

Experiment 1: Simulation-Based Sensitivity analysis

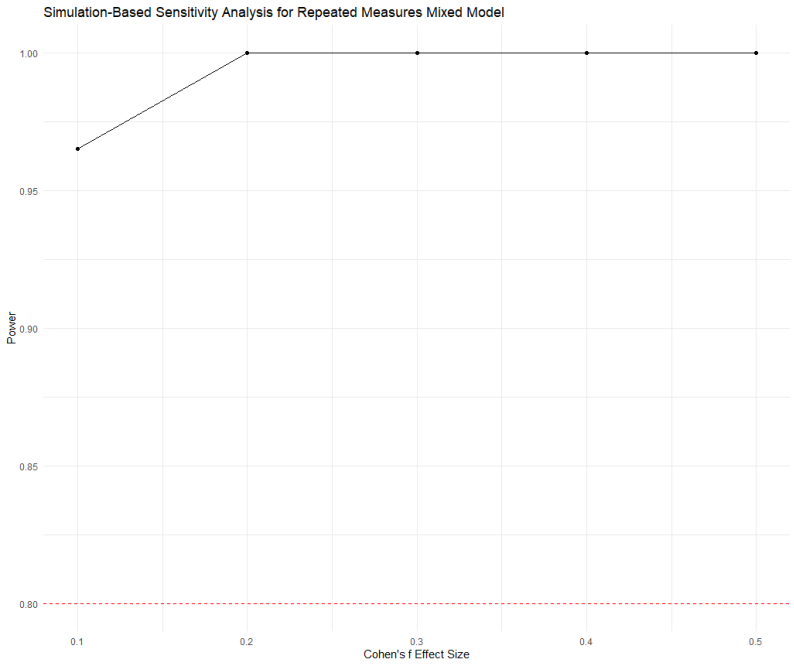


Figure S2

*Experiment 2: S*imulation *modelled* repeated measures with a within-subject correlation.

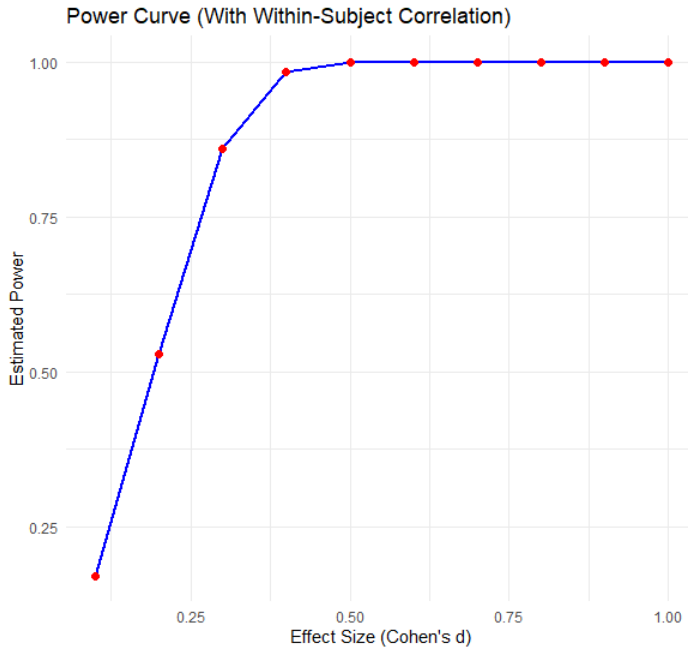


Figure S3

Density Plots of Greek and English pre-test scores (English pre-test 1 vs English pre-test 2 & Greek pre-test 1 vs Greek pre-test 2).