UNIVERSITY OF PADOVA

Department of Developmental Psychology and Socialization

Laurea Magistrale in
Developmental and Educational Psychology

Final Dissertation

The Trajectory of Moral Judgments in Foreign Language: A MouseTracker Investigation

Supervisor:
Professor Eduardo Navarrete Sanchez

Co-supervisor:
Zhimin Hu

Candidate: Shirin Kiasoroush

Matricola Number: 2042054

Academic year 2022/2023

Contents

Preface

As the people of the world are becoming increasingly bilingual or multilingual, and as the portion of immigrants is growing in various countries, studying the effects of bilingualism is becoming a relevant research topic in psychological sciences. Knowing whether using a foreign language affects our behavior, emotions, and moral judgments is becoming more attractive in the field of psychology every day. This knowledge is no longer just interesting but also important and necessary to understand the underlying psychological mechanisms in the decision-making process of many people all around the world. As an international student in Italy, I need to speak three different languages in different contexts every day. I use Farsi -my native language- to talk to my Iranian friends and family, read the news, listen to songs, read books, etc. I use English to study, speak to my professors and international friends, watch movies, and do my job. I also use Italian to communicate with Italian people. I have found differences in how I feel and think in each language. As a psychology student, I started monitoring myself and my friends to recognize what changes occur to us by changing the language and how we feel about these changes. After researching and reading scientific sources, I decided to continue my academic efforts on this path. That is why this thesis focuses on the effects of using a foreign language on moral decisions using the MouseTracker technique. I believe this topic sheds light on different aspects of our cognitive and emotional mechanisms and also has practical benefits for psychologists working with immigrants and bilinguals in various cross-cultural contexts.

Abstract

This thesis explores the complexity of the foreign language effect (FLE) on moral decision-making processes using the MouseTracker technique. Building on previous research, we focus on native Italian speakers to understand how a foreign language (English) influences various aspects of moral judgment in bilinguals. This study utilizes the innovative MouseTracker software to capture real-time cognitive processes, providing a comprehensive view of decision-making dynamics. Our experimental design includes variables of Language (Native Language versus Foreign Language), Risk Involvement (self-involved versus other-involved), and Type of Dilemma (instrumental versus incidental) alongside their interactions.

The findings indicate that while foreign language plays a moderating role in moral decision-making, it does not mainly affect and increase utilitarian responses, challenging some existing literature. Temporal dimensions of decision-making and mouse trajectory metrics further reveal patterns of cognitive conflict and decision-making complexity, offering insights into the underlying mechanisms of moral judgments in a bilingual context. This research contributes to the growing field of linguistic psychology, providing a detailed and valuable understanding of how language shapes our moral judgment and decision-making processes.

Chapter 1
The Domains of the Foreign Language Effect

Imagine that you are facing a crucial situation in your professional life and need to make an important decision. For instance, imagine that due to your job, you have access to confidential information about a stock, and your close friend is willing to invest in that stock. You know that he will lose all his money if he does so. Should you give confidential information to your friend and prevent him from investing all his money in a stock you know will have a huge loss, or should you be loyal to your professional morality? Or imagine dealing with a health disease and, after having visited two doctors who gave you two different treatment programs, you have to choose which one to follow. One might think these decisions are so crucial that aspects like the frame or the language in which the information is given would not affect the decision-making process. However, recent research on decision-making has demonstrated that language context, the language –foreign or native- in which the information is presented may determine the final decision. This phenomenon is called the foreign language effect (FLE). FLE refers to the differences in our preferences, decisions, and judgments while using a foreign language. In the current study, we explore this effect by recruiting our sample from bilinguals whose native language (NL) is Italian and English is their foreign language (FL).

1.1. Bilingualism

Bilingual individuals have the ability to produce and understand two different languages. They may use different languages in different contexts, such as working, studying, or adjusting their relationships with others. These activities require comprehending, thinking, producing, and making decisions in both NL and FL. However, this is not just an ability to understand and use different languages; it also influences their decisions and judgments on various levels. To explore the language frame effect in decision-making, we focus on a specific topic of decision-making: moral judgment in moral dilemmas.

Bilingualism or multilingualism can be defined as "the coexistence of more than one language system within an individual, as contrasted to monolingualism." (Hakuta, 2009, p. 173). However, offering a comprehensive definition of bilingualism is challenging because this phenomenon includes a wide range of people who use two languages. These people vary in the number of languages they speak (two, three, or more), the age at which they acquired the languages, the fluency level in each language, the frequency of using each language, and the context they use each language in (see Green & Abutalebi, 2013). As we will see below, the type of bilingualism is a critical factor as it modulates the FLE. That is, the influence of the FL in the decision is observed only (or mainly) in unbalanced bilinguals. Unbalanced bilinguals, contrarily to balanced bilinguals, are not equally proficient in two languages and are more proficient in one language than the other (Harris, Cullum & Puente, 1995; Rosselli, Ardila, Lalwani & Vélez-Uribe, 2016).

To explore the FLE on moral dilemmas in the current study, we focus on those bilingual individuals who are not balanced. Our sample lives in a monolingual context corresponding to their NL (Italian), and they have acquired the FL (English) mainly at school.

1.2. The Foreign Language Effect in The Literature

The foreign language effect (FLE) refers to the observation that decisions, emotions, and preferences can differ depending on employing the NL or FL. A growing body of research focuses on this effect, its consequences, and its causes. In the following, we review five research studies: three on the FLE on cognition and two on the FLE on emotion. Then, we explain the domains and proposed explanations of the FLE in the literature.

Keysar, Hayakawa, and Yu An (2012).

Keysar et al. (2012) did the first study reporting the interaction between language and decision-making. In a series of experiments, Keysar and colleagues showed that the *framing effect* and *loss aversion* biases are reduced when using the FL. The framing effect is a type of cognitive bias referring to the observation that the way information and options are presented to us – whether they are framed positively or negatively- can influence our decisions (Kahneman & Tversky, 1979). For example, people perceive it more positively if a product is labeled 80% fat-free compared to when it is labeled 20% fat. The actual percentage of the fat is equal on both labels, yet people tend to buy the first product more because the information is framed more positively. Loss aversion is another cognitive bias indicating that people

tend to outweigh the negative outcome of a potential loss over a potential gain. For instance, people tend to refuse a bet offering them an even chance of winning 12 $ or losing 10 $. Here, the fear of loss overweighs the possibility of winning and makes people reject the bet (Kahneman & Tversky, 1979).

Keysar and colleagues used a modified version of the Asian disease scenario and a betting scenario to investigate the FLE on cognitive biases. In the Asian disease scenario (Tversky & Kahneman, 1981), a hypothetical fatal disease afflicts 600,000 people for which two possible treatments are available. The possible treatments are provided either in the gain or loss frame. The gain frame indicates: "If you choose Medicine A, 200,000 people will be saved. If you choose Medicine B, there is a 33.3% chance that 600,000 people will be saved and a 66.6% chance that no one will be saved". Whereas the loss frame states: "If you choose Medicine A, 400,000 people will die. If you choose Medicine B, there is a 33.3% chance that no one will die and a 66.6% chance that 600,000 people will die". Critically, the outcomes of the treatments are identical, yet they differ in the frames in which they are presented.

The sample was recruited from different bilingual populations. For each NL, the experiment was done in two groups; one performed the task in their NL and another in their FL. Native English participants were tested in English as their NL or in Japanese, French, or Spanish as their FL. Native Korean participants were tested in Korean as their NL or English as their FL. The results indicated a bias reduction for the groups performing the task in the FL compared to those performing in the NL, implying that the FL reduces the loss aversion and framing bias. Another critical

finding of this study is that bias reduction was observed in all bilingual populations, indicating that the FLE is present beyond a particular language.

In the second scenario, the betting scenario, each participant was presented with equal odds bets with positive expected values. Half the bets had high stakes (e.g., lose 119,000 ₩ or win 170,000 ₩), and half had low stakes (e.g., lose 200 ₩ or win 500 ₩). Participants were native Korean speakers who participated in Korean language (NL) or English (FL). On average, participants took more bets in English than in Korean, which replicated the findings of the Asian disease scenario experiment, indicating that people are less cognitively biased in their FL than in their NL.

Gao, Zika, Rogers, and Thierry (2015).

In line with Keysar and colleagues' findings, other studies have also reported a reduction in different cognitive biases while using an FL. One of these biases is the *hot-hand fallacy* studied by Gao et al. (2015). The hot-hand fallacy refers to people's tendency to interpret a positive outcome in a trial (e.g., winning) as a predictor of a positive outcome in the upcoming trial. For instance, basketball fans assume that the chance for a player to hit a basket is greater if it follows a hit rather than following a miss. (Gilovich, Vallone & Tversky, 1985). In other words, people assume the good or bad outcomes are correlated with the subsequent trials and influenced by them. Gao et al. (2015) studied the FLE on this bias. They offered Chinese-English bilinguals a series of gambles with even probabilities of winning and losing, either in English (FL) or Chinese (NL). In English trials, fewer gambles

after positive game outcomes were observed, indicating the FLE in reducing cognitive bias.

Díaz-Lago and Matute (2019).

Díaz-Lago and Matute conducted experiments on native English and native Spanish speakers to study the FLE on the *causality bias* (or illusion of causality). This bias refers to developing the belief of a causal relation between two events while there is no actual causality between them.

To detect the FLE on causality bias, they gave the participants some information in different trials, in their NL or FL. Participants had to decide whether the trials had a causal relation. Results showed that participants who performed the task in their FL had more accuracy in detecting that the two presented events were causally unrelated. Hence, the FL had reduced the bias in their judgment.

Ivaz, Costa, and Duñabeitia (2016).

For most bilinguals, NL is learned in contact with others and in emotionally rich contexts, while FL is acquired in academic, formal environments. Therefore, FL can be aligned with less emotionality. Ivaz et al. (2016) conducted their study to explore whether this difference in emotionality in NL and FL can be proven by observing the FLE on reducing self-bias in response to self-related linguistic stimuli. Self-bias leads to faster and more accurate responses to the stimuli that are self-related compared to the self-unrelated ones.

Ivaz and colleagues presented a perceptual matching task to the participants. Participants had to quickly associate three simple geometric shapes (circle, square, and triangle) with three labels: you, friend, and other. The matching pairs were "circle + you", "square + friend", and "triangle + other". Any other form of matching should have been considered mismatched. The authors assumed that if the FL reduces emotionality, self-prioritizing or self-bias would be reduced in FL. In the first experiment, Native Spanish speakers with high English proficiency did the task in two groups: one in Spanish and the other in English. In the second experiment, Native English speakers did the task in English. A robust ELE of more self-bias was observed in participants responding in their NL (both Spanish and English) compared to the ones responding in the FL (native Spanish speakers in English). This observation indicated an emotional distance associated with the FL, affecting quick and automatic emotional processing.

Dewaele (2008).

Dewaele Studied the emotionality aligned with the phrase "I love you" in several native and foreign languages. The participants were 1459 adults, including bilinguals, trilinguals, quadrilinguals, and pentalinguals, from 77 NLs. They joined the experiment online and answered these open-ended questions: "Does the phrase "I love you" have the same emotional weight for you in your different languages? Which language does it feel strongest in?". They also filled out self-reports to declare their dominant language, FL(s) age and context of acquisition, frequency, and network of using FL(s), and oral proficiency in FL(s) plus a trait emotional intelligence questionnaire—short form (TEIQue-SF).

The majority of participants reported that they felt "I love you" stronger in their NL. This finding was not associated with socio-biographical variables (e.g., gender and education) nor by trait of emotional intelligence. However, it was related to learning history, recent usage, and self-perceived proficiency of the FL(s). In the sense that these factors could lead to more or equal emotionality, align with the phrase "I love you" in the FL compared to the NL.

This study is not an experimental study like the other studies we mention in our work, and it should be considered that the self-report measurements may have some biases. Nevertheless, the large number of participants from various NLs and FLs, along with the robustness of observing the FLE on emotionality distance, makes this study valuable evidence of emotionality reduction in FL.

1.3. Explanation of the FLE

The theoretical frame that has been proposed to account for the bias reduction underlying these studies is the dual-process model of Kahneman (2003). According to this model, there are two different thought processes or decision-making models; the first is a more automatic, emotional, intuitive, and unconscious process, whereas the second is more of an explicit, cognitive, systematic, and conscious process. There are at least three reasons why FL and NL may affect these two systems. The first reason refers to the claim that processing information (i.e., linguistic input) in FL is less automatic than in NL and requires more cognitive effort. The second reason is

that FL has less emotional engagement in comparison with NL (e.g., Costa, Foucart, Hayakawa, Aparici, Apesteguia, Heafner & Keysar, 2014; Incera, Tuft, Fernandes & McLennan, 2020). There is also a third account for the FLE: the notion of moral dilemmas and our social decisions are determined by the social norms we have acquired. Commonly, social norms acquisition for most people occurs in the NL and not in the FL. Below, we detail these three accounts in more detail.

Cognitive effort. One explanation for the FLE is the cognitive load, which refers to the amount of working memory (and cognitive effort) resources used. To process and solve problems, humans need to sort and hold a large number of items in their short-term memory, which can lead to an excessive cognitive load (Sweller, 1988). For bilinguals who are less proficient in their FL than their NL, a cognitive effort is needed to inhibit and control the NL to use the FL (Costa, Hernández & Sebastián-Gallés, 2008). For them, the FL usage is more demanding than the usage of NL. Therefore, using an FL brings more cognitive effort (Keysar et al., 2012). The cognitive effort results in cognitive load, leading to relying more on the second system process (Del Maschio, Crespi, Peressotti, Abutalebi & Sulpizio, 2022). Less fluency can also make the decision-making process slower, moving individuals toward the second system, which is more careful and deliberative (Costa et al., 2014).

Emotional processing. The other aspect of relying on the second process model is emotionality reduction. Since the NL is acquired early in life and usually in the family context, it is aligned with a wide range of emotionally engaged experiences. This leads to encoding emotions with native words in the memory. In contrast, an FL is learned chiefly later in life, mostly in a classroom environment

with a less emotionally engaged atmosphere than the family context. Therefore, it can bring less immersion than the NL (Caldwell-Harris, 2014). Bilinguals have shown reduced emotionality in different studies, such as the *emotional Stroop task* in their FL. In the emotional Stroop task, participants are presented with emotional words and neutral words in different colors and are asked to determine the color of the words. Sutton, Altarriba, Gianico, and Basnight-Brown (2007) used the emotional Stroop task to measure the FLE. Their participants were native Spanish speakers and were presented with emotional and neutral words in both Spanish and English separately. Participants showed longer reaction time to emotional Spanish words than emotional English words. This result indicates the reduction of emotional reaction while using an FL, as well as relying more on the second process system. Another support for emotionality reduction in FL is the studies on the difference between reactions to taboo words in NL and FL. In every society, some words are considered inappropriate in public situations. Using these words would cause negative emotions such as anxiety, embarrassment, or shame for the people. Different societies have a wide range of taboo words that "can vary widely: sex, death, illness, excretion, bodily functions, religious matters, the supernatural." (Gao, 2013, p. 1). Studies have shown that bilinguals have a delayed and less intense reaction to taboo words in FL, which supports the idea of an emotional reduction in FL. Harris**,** Ayçıçeğı & Gleason (2003) studied the difference in emotion reactivity to different categories of words in Turkish–English bilinguals in their NL (Turkish) and FL (English). They were visually or auditory exposed to taboo and non-taboo words. During this process, their skin conductance response (SCR) was monitored. Then, they rated blocks of words for the emotional intensity of the word, followed by a recall task. The results of all the measurements -SCR, emotionality rating, and

the recall task- showed greater reactivity to taboo words in both languages with a slightly stronger SCR in the NL, especially in the auditory modality.

Social norms account. As we saw until now, researchers mainly focus on cognitive and emotional components to describe the FLE. However, this effect has also been explained from another aspect, which refers to reducing social norms while using the FL. This concept has especially been explored in moral judgments (situations where an individual must make a moral decision between two or more available options). Geipel, Hadjichristidis, and Surian (2015) studied the FLE on moral judgments in different types of moral violations. They also studied whether FLE influences moral judgment through the attenuation of emotions. Their participants -native German speakers and native Italian speakers- had to judge the degree of the wrongness of some moral scenarios on a scale ranging from 0 (perfectly okay) to 9 (extremely wrong). Native German speakers were tested in English, Italian, or German, and Native Italian speakers were tested in English or Italian. After each scenario, participants rated their emotionality and indicated how upset, worried, disgusted, sad, and angry they felt while reading each scenario. The results indicated that using an FL elicited less harsh moral judgments in all types of moral violence. Meanwhile, according to the emotional rating scale, limited support for the mediation of attenuation of emotions in FLE was observed. Since this attenuation was shown in some of the scenarios and not all of them, Geipel and colleagues propose that the FLE on moral judgments appears due to the "reduction of the mental accessibility of moral and social rules.". This effect emerges because the social norms are learned and encoded in the mind, directly and indirectly, in the plain of the NL. Therefore, they can be recalled and accessed more easily while using NL than FL.

In congruence with Geipel and colleagues' findings (2015), Sulpizio, Toti, Del Maschio, Costa, Fedeli, Job, and Abutalebi (2019) also referred to the FLE on social-norm processing as part of their explanation for their findings of FLE. They used fMRI to explore the FLE in processing taboo and non-taboo words. To explore this effect, they asked Italian monolinguals (in Italian) and highly proficient Italian-English bilinguals (in Italian and English) to differentiate between taboo and non-taboo words while using fMRI. Their brain imaging results indicate that taboo words and socio-pragmatic knowledge are automatically conveyed in NL. At the same time, it requires more effort and engagement of additional brain structures to represent and evaluate the taboo words and socio-pragmatic knowledge in the FL. It is noteworthy that contrary to Geipel et al. (2015), Sulpizio et al. (2019) do not refuse the role of emotionality reduction in FL. To describe this effect, they indicate less automatized access to socio-pragmatic and emotional information in an FL.

We have reviewed the FLE and its various aspects and explanations in this chapter. In order to investigate this effect on moral judgment, we will look into the phenomenon of moral dilemmas in the next chapter.

Chapter 2

Moral Dilemma, Its Types, and Possible Responses

As explained in the previous chapter, the FLE is observed in different domains, especially decision-making. Moral dilemmas are one of the decision-making situations in which this effect can show significant consequences. A *moral dilemma* is a situation in which an individual has to make a moral decision between two or more available options. We may face these dilemmas on different levels on a daily basis. For instance, is it ethical to tell a white lie while our coworker had made an innocent mistake, and telling the truth would lead to firing him? Is it acceptable to pass the red light and put other cars in danger when we have a passenger who must be taken to the hospital as soon as possible? Choosing each one of the options would violate at least one crucial moral concern (Kvalnes, 2019). In other words, in moral dilemmas, we have to weigh different moral principles against each other and decide which is more moral and right to be applied in the given situation. Whichever one of the choices we choose, we can define it based on specific moral rules to justify breaking some others. In the following, we will explain this phenomenon, its types, and possible responses in more detail.

Footbridge and Trolley Dilemmas. In various fields of study, the reasons and mechanisms behind the choices we make in moral decisions have been investigated by applying different moral scenarios. Two of the most well-known dilemma scenarios are the "*trolley*" and "*footbridge*" dilemmas, based on which many other moral scenarios have been developed. In the trolley scenario, a runaway tram or

trolley is about to hit and kill five people on its way. The driver can save these five people only by hitting a switch and turning the trolley onto a sidetrack where it would kill only one person. Here, the question is which of these two options is more ethical: to hit the switch and sacrifice one person's life to save five other people or let the tram continue its path and kill five other people (Philippa Foot, 1967). In the footbridge scenario, a person standing on a footbridge sees a trolley hurtling down the track, and it is about to kill five workers. The only way to stop it and save the workers is to throw an overweight man standing on the footbridge into the path, which will cause his death. Also, in this case, the challenge is to decide which is more ethical: to push the overweight man and save five lives or to let the trolley continue on its path and kill five people. (Judith Jarvis Thomson, 1985). The main difference between the trolley and footbridge dilemma is that in the trolley scenario, the lives of five people can be saved by pushing a button, whereas in the footbridge scenario, the overweight man must be pushed off the bridge by someone’s hands.

Personal and *impersonal dilemmas.* The footbridge and trolley dilemmas can also be put down as *personal* and *impersonal* dilemmas, respectively. In personal or footbridge-like dilemmas, the force that impacts the other person and causes the death is direct and generated by the agent's muscles, like pushing the man off the bridge by hand. Therefore, the act in personal dilemmas cannot be mediated by other things, even mechanisms generated by the agent. Conversely, in impersonal or trolley-like dilemmas, there is serious bodily harm to a particular person, causing the death, but the harm is not the result of a direct act; it happens through something like pushing a button. This difference can result in more emotional engagement in footbridge-like dilemmas (Greene, Sommerville, Nystrom, Darley & Cohen, 2001).

Utilitarian and deontological choices. The possible responses to moral dilemmas can be divided into utilitarian and deontological. The utilitarian responses determine the morality of an action by its consequences, e.g., saving "five" lives is more moral than saving "one" life. Whereas the deontological responses rely more on the intrinsic nature of the action, e.g., killing a person is immoral even to save more people's lives (Conway & Gawronski, 2013). Greene, Morelli, Lowenberg, Nystrom, and Cohen (2008) used the dual-process model to distinguish between these two moral reasoning processes and describe the moral judgment process more precisely.

According to Greene and colleagues (2008), utilitarian judgments stem from brain regions responsible for analytical thoughts, whereas deontological moral judgments stem from regions responsible for emotions. As the authors state, "dual-process theory specifies that automatic emotional responses incline people to disapprove of pushing the man off the footbridge, while controlled cognitive processes incline people to approve of this action" (Greene et al., 2008, p. 1145). Using fMRI, Greene et al. observed that when people face footbridge-like dilemmas, it takes longer to judge moral violations as appropriate than inappropriate. At the same time, this reaction time difference was not present when facing trolley-like dilemmas (Greene et al., 2001; Greene, Nystrom, Engell, Darley & Cohen, 2004).

In the FLE section, we have discussed the dual system theory of Kahneman, according to which there are two systems for making decisions and judgments; the first is a fast, unconscious, and effortless effective system competing with the second one, which is a slow, conscious, and effortful cognitive system. Based on this theory,

Greene et al. demonstrated that cognitive load manipulation interferes with reaction time in utilitarian judgment, indicating that a cognitive load increases utilitarian reaction time while not influencing non-utilitarian judgment (Greene et al., 2008). In accordance with this finding, Koenigs, Young, Adolphs, Tranel, Cushman, Hauser, and Damasio (2007) showed that patients with focal bilateral damage to the ventromedial prefrontal cortex, an area necessary for generating emotions, choose more utilitarian options abnormally.

In the two main topics above (FLE and moral dilemmas), we discussed how using an FL can affect our decision-making and how moral dilemmas are a context of making a decision. Therefore, decision-making is the common point of these two psychological topics together. Past researchers have studied the decision-making in this situation by exploring how the situation of the dilemma (e.g., personal or impersonal) modulated the percentage of deontological and utilitarian decisions. In addition, there is also research exploring moral dilemmas by analyzing reaction times, that is, the time it takes participants to make a deontological or heuristic decision. The current research aims to explore moral dilemma decisions by exploring a relatively recent methodology, MouseTracker. Before discussing this method, we present in Chapter 3 the research that has already explored the interaction between FL and moral decisions. Then, in Chapter 4, we introduce the MouseTracker technique and argue why this technique offers fine-grained information that can be useful in describing how decision-making is resolved in moral dilemmas. Then, in Chapter 5, we introduce the main aim of the research: to use MouseTracker to explore the differences in the underpinning mechanisms of decision-making in moral dilemma resolutions in NL and FL.

Chapter 3
The Foreign Language Effect on Moral Dilemmas in the Literature

In the previous chapters, we discussed the domains of the FLE and the concept of a moral dilemma, its types, and the factors through which an FL may influence our choices while facing a moral dilemma. This chapter joins these two main topics and reviews five influential experimental studies and two recent meta-analyses investigating the FLE on moral dilemmas. These studies demonstrate that although we may assume that our moral beliefs and judgments have such profound roots and logic that the way a moral situation is presented cannot affect our decision, it is not the case; the language (NL or FL) in which we confront a moral situation can influence the way we perceive that situation and the final decision we make.

3.1. Costa, Foucart, Hayakawa, Aparici, Apesteguia, Heafner & Keysar (2014)

Costa et al. studied the effects of using the FL on the footbridge and trolley dilemmas in two sets of experiments. In experiment 1, they presented the footbridge dilemma to the participants in several native and foreign languages in different countries: native English speakers were experimented with either English, Spanish, French, or Hebrew; native Korean speakers were experimented with either Korean or English, and native Spanish speakers were experimented with either Spanish or Hebrew. Participants received a written version of the dilemmas and their possible options, plus a cartoon depiction of the dilemma's scene. In this experiment, more participants chose the utilitarian choice (killing one person to save five lives) when experimenting with their FL than their NL. This observation indicates the reduced

emotional resonance aligned with the FL, resulting in less emotional aversion regarding pushing the overweighed man, hence making more utilitarian choices.

To control other possible explanations rather than the reduced emotionality in the FL (e.g., random responding or cultural norms), Costa et al. ran experiment 2 in which all participants were presented with the written form of both the footbridge and trolley dilemmas in their NL or FL. The native Spanish speakers were experimented with either Spanish or English and native English speakers were experimented with either English or Spanish. The findings of experiment 2, in line with experiment 1, showed a higher choice of pushing the man and causing his death in favor of saving five other people while being experimented on in the FL compared to the NL.

Another noteworthy aspect of Costa and colleagues' study is the consideration of *language proficiency* on the FLE. They divided the participants into two proficiency levels, above-average or below-average, based on self-rated proficiency level. An increase in utilitarian choices was observed in an FL for both proficiency levels, according to a post-hoc analysis. Meanwhile, this increase was more significant for the below-average group than for the above-average. This result reveals an essential aspect of the FLE; as the proficiency level of the FL increases, it becomes more emotionally grounded (Dewaele, 2004); therefore, the proficient bilinguals may show less difference between the conditions using their NL and FL.

3.2 Cipolletti, McFarlane & Weissglass (2016)

Cipolletti and colleagues studied the FLE on moral decisions regarding the dual-process model. To evaluate this effect, they provided two experimental cases of "the button case" and "the bridge case", which are trolley-like and footbridge-like dilemmas, respectively. In both cases, an unoccupied runaway train is hurtling down the tracks and is about to hit and kill five innocent people. In the button case, there is a button that would lead the train to a side track on which there is just one innocent person who would get hit and die. The case finishes with this question: "Morally speaking, should you push the button to direct the train to the sidetrack?". The scenario of the bridge case is the same, but the difference is that the only way to save the five people is to push an innocent person you do not know onto the bridge and cause his death. This scenario ends with this question: "Morally speaking, should you push this person onto the tracks below?". The participants could answer the dilemma by choosing the options "yes" and "no".

Participants of this study were native English speakers and native Spanish speakers. Each participant received one of the cases in the form of a questionnaire in either their NL or FL (English for native Spanish speakers and Spanish for native English speakers). It is worth noting that the sample size of native Spanish speakers was smaller than the native English speakers, and the number of native Spanish speakers who did the task in English was zero for the Button case and one for the Bridge case. The majority of the participants doing the task in their NL chose the utilitarian option of pushing the button to save five lives but did not choose the utilitarian choice of pushing the stranger off on the track to save more lives. For the

FL group, the ones answering the button case acted like the ones answering in their NL; the majority chose to push the button to sacrifice one life to save more lives. However, there was a difference in the bridge case between the NL and FL groups; the ones answered in their FL chose more utilitarian choices of pushing the man off the bridge to save the lives of five other people. This difference between the results of the Bridge case in NL and FL led the authors to conclude that deciding on the moral dilemmas while using FL relies on the second system activation and systematic reasoning that increased more utilitarian choices.

3.3 Corey, Hayakawa, Foucart, Aparici, Botella, Costa, and Keysar (2017)

In an influential and comprehensive study, Corey et al. replicated the findings of Costa and colleagues (2014) while measuring and controlling other possible explanations and factors for the FLE. They conducted nine experiments with 2,000 native Spanish speakers, and in all experiments, they did the tasks either in Spanish as the NL or in English as the FL.

In Experiment 1a, they applied the materials of Costa et al.'s study (2014) by presenting the trolley and footbridge dilemmas to the participants. Their results indicated a significant effect of the FL on choosing more utilitarian choices (killing one person in favor of saving more lives) in both the trolley and footbridge dilemma. They tested this finding on another set of dilemmas in Experiment 1b to evaluate the generalizability of the last experiment's finding. To do this, they presented the participants with the *Hospital* (indirect, Impersonal, and trolley-like) and the *Terrorist* (direct, personal, and footbridge-like) dilemmas. In the Hospital

Dilemma (Adapted from Thomson, 1985), a fire causes smoke to enter a hospital through the ventilation system. There are five patients in one room and one patient in another room. If nothing is done, the smoke will kill the five patients. There is a button that can divert the smoke to the room with only one patient, saving five patients but the death of a single patient. The participants had to decide whether they, as a person working in the hospital, would push the button or not. In the Terrorist Dilemma (Adapted from Greene et al., 2001), the leader of a terrorist group capturing six tourists offers a choice: If you choose a tourist and kill him, the other five will be set free. If not, the terrorists will kill five tourists and set one free. The participants must decide whether they shot someone dead to save another five people. Again, their results showed the FL's significant effect on killing one person to save more people in the terrorist dilemma. In contrast, this difference was not significant in the hospital dilemma.

There is an argument that the reason behind observing the effects of an FL is not the result of the FL per se, and the emotion reduction in a non-native language can be due to the *cognitive control* occurring while *language-switching* (Oganian, Korn, & Heekeren, 2016). Therefore, as Greene et al. (2008) demonstrated, cognitive control and cognitive load increase the probability of choosing utilitarian options. To test this argument, in Experiment 2a, Corey et al. presented the participants with the trolley dilemma in Spanish or English, followed by the footbridge dilemma in the other language. Again, although all participants switched languages, more utilitarian choices were observed in the footbridge dilemma while experimenting with FL than in the NL. At the same time, no effect of the FL was observed in the trolley dilemma.

In Experiment 2b, Corey et al. addressed another possibility for the observed results: the *social norms of in-group and out-group membership* influencing the choices in moral dilemmas. To assess this possibility, they presented each participant with the in-group or out-group versions of the trolley or footbridge dilemma. The in-group versions were precisely like the ones used in experiments 1a and 2a, with a change at the first line that indicates the five people are Spanish for the in-group version and American for the out-group version. However, there was no effect of the in-group or our-group condition, and the participants chose to kill one person to save five people in the footbridge dilemma while using the English language. Also, there was no difference between the choices in the trolley dilemma, neither in the in-group and out-group condition nor in the use of the FL or NL.

In experiment 3a, Corey and colleagues evaluated whether a less aversive action in the footbridge dilemma would demonstrate the same results. They provided the participants with the "button" dilemma as an adapted version of the footbridge dilemma. Instead of pushing the man directly, a button must be pushed to throw him on the track. The consequences were the same as the original footbridge dilemma; the man would be pushed off and killed, and the other five people would be saved. Participants were tested by this button dilemma and the trolley dilemma. If the FLE stems from reduced action aversion, it should be absent when the scenario includes a less aversive action (pushing a button instead of pushing the man directly). Still, using the FL resulted in more utilitarian choices (pushing the button). Meanwhile, there was no language effect observed in the trolley dilemma.

The effect of focusing on the *consequence* instead of the action was assessed in Experiment 3b. To do so, the questions at the end of the original footbridge and trolley dilemmas were changed from "Would you push the man?" and " Would you change the track?" to "Would you let five people die?". Even so, there was a significant ELE in the consequence-focused version of the footbridge dilemma, and the participants who did the dilemma in English chose more utilitarian choices. However, the effect was not as strong as the results of the original version of the dilemmas (Experiment 1a). Moreover, no such effect was found for the consequence switch dilemma.

In Experiment 3c, Corey and colleagues went one step further toward assessing the influence of the consequence by evaluating the *trade-off between the means and consequences*. This evaluation was conducted by changing the final questions at the trolley and footbridge dilemmas to "Would you let five people die by not changing the track?" and "Would you let five people die by not pushing him?" respectively. This change was done to study the possibility that using an FL would prompt more focus on the consequence than the trade-off between the means and consequences. In contrast with the last experiments, no significant effect was observed between the ones who used English and those who used Spanish neither in the consequence action footbridge dilemma nor the consequence action trolley dilemma. Experiments 3b and 3c indicate that the wording of the questions can modulate the FLE in dilemmas. In comparison with the original versions of the dilemmas (Experiment 1a), the FLE was *weaker* when the consequence was more explicit (Experiment 3b) and *null* when the trade-off between the means and consequences was explicit (Experiment 3c).

Experiments 3d and 3e examined whether, in the footbridge dilemma, the *change in consequence* of pushing the man from his *death* to him *getting paralyzed* or *injured* changes the participants' preferences. Therefore, in Experiment 3d, some participants received the original version of the trolley dilemma, whereas some received the modified "disabled footbridge" dilemma. This modified version had the same scenario as the original footbridge dilemma, but the consequence of pushing the man was not his death. Instead, pushing the man would paralyze him and make him unable to walk for the rest of his life. In Experiment 3e, the trolley dilemma and another modified version of the footbridge dilemma were used; the modified "injured footbridge" dilemma which indicated that by pushing the man on the truck, he would be seriously injured but not paralyzed or dead. In both experiments, participants who did the task in English made more utilitarian choices of paralyzing (in experiment 3d) and injuring (in experiment 3e) the fat man to save five people than those examined in Spanish (NL). However, these effects were not statistically significant, and no effect of language was observed between the ones who experimented with the trolley dilemma in their NL and FL. The two latter experiences demonstrate that the FLE can be absent when the consequence of saving more people is less severe and persistent than the death consequence.

3. 4 Hayakawa, Tannenbaum, Costa, Corey & Keysar (2017)

The prior examined research has demonstrated that employing the FL tends to result in more utilitarian choices. Hayakawa and his team have presented inquiries concerning these findings, seeking to uncover whether the noted effect stems from a reduction in emotional responses during FL use, a reduction in deontological choices

(as suggested by Costa et al., 2014), or is a consequence of system two engaging leading to an increase in utilitarian choices (as proposed by Cipolletti et al., 2016).

To reveal where the FLE on moral dilemmas stems from, Hayakawa et al. conducted six experiments in different native and foreign languages by adopting the dual-process framework (Stanovich & West, 2000) to determine the underlying mechanism of the Moral Foreign Language Effect (MFLE). Among the six experiments, the first two were done on original dilemmas from Conway and Gawronski (2013), and native German speakers did the task in either German (NL) or English (FL), whereas native English speakers participated in English (NL), or Spanish (FL). Experiments 3 to 6 were done on Revised dilemmas from Conway and Rosas (2017). The participants did the tasks as follows: native Spanish speakers in either Spanish (NL) or English (FL), Native German speakers in either German (NL) or English (FL), and native English speakers in either English (NL) or German (FL). All six experiments were conducted online except for Experiment 3, which was done in a laboratory setting, and each participant answered 20 dilemmas.

The dilemmas had diversity in three aspects. First, they differ in their emphasized point in the question at the end of the dilemmas. Experiments 1 and 2 were about *judgment* (e.g., "Is it appropriate to push the man off the bridge?"), in Experiments 3 and 4, *judgments with consequences* were highlighted (e.g., "Is it morally correct to push the man off the bridge to save five people, even though the man would die?"), Moreover, Experiments 5 and 6 were about *choice* (e.g., "Would you push the man off the bridge to save five people?"). All questions offered three options of "yes," "no," and "I do not understand". Second, some experiments

(Experiments 1 to 4) asked participants whether the given action was *appropriate*, and some (Experiments 5 and 6) asked whether they were *willing to perform the action* themselves. Third, ten dilemmas were congruent for each participant, and ten were incongruent. Dilemmas like the traditional Trolley and Footbridge dilemma are incongruent because they contain a conflict between deontological and utilitarian concerns. Whereas in the congruent dilemmas, the deontological and utilitarian choices are in agreement. For example, if the dilemma states that killing one person would prevent five people from getting mildly injured, it would be a congruent dilemma; none of the deontological and utilitarian concerns lead to killing that person.

In order to reach measures for deontological and utilitarian responding, Hayakawa and colleagues used two formulas to compare the response rates of the congruent and incongruent dilemmas. They calculated the utilitarianism parameter (U) by comparing the "no" responses proportion between congruent and incongruent trials. Higher U scores indicate that participants found harmful actions unacceptable when they did not maximize welfare (congruent trials) but acceptable when they did (incongruent trials). Then, they calculated the deontological considerations (D) by determining the proportion of responses not driven by utilitarianism. Higher D scores indicate stronger deontological responses. Lower D scores for participants using the FL than those using the NL were observed in all the experiments. However, the results failed to support a higher U score among the ones responding in their FL. Even in three experiments, a decrease was observed in the U parameter for the ones who participated in their FL. These results provide clear evidence for the claim that

the MFLE occurs due to the blunt of the deontological response rather than an increase in utilitarian choices per se.

3. 5 Miozzoa, Navarrete, Ongisc, Mellob, Girottod & Peressotti (2020)

Previous studies on bilingualism and FLE have focused on a specific type of bilingualism in which one language is considered the NL and one the FL. In these cases, the NL is the first language learned in life, mostly in the family environment and through interaction with others. In contrast, the FL language is acquired later, mainly at school or in a classroom setting. This group of bilinguals usually has less proficiency in their FL than their NL and shows an effect of language while making decisions in each of these languages. Miozzoa and colleagues assessed whether the FLE on emotionality and decision-making that had been proven in the previous studies could be extended to another type of bilinguals: the ones who have two native languages, learned both of them early in life, use both regularly, and are proficient in both.

To answer the question above, Miozzoa and colleagues studied Italian-Venetian and Italian-Bergamasque bilingual participants in four studies. Venetian is a Romance language spoken in the northeast of Italy, mainly in the region of Veneto, and the Bergamasque dialect is the western variant of the Lombard language, spoken in the city of Bergamo and the nearby area. Italian is the formal and official language in both these areas, as in other parts of Italy. Inversely, Venetian and Bergamasque are local, informal languages used domestically to communicate with family members, friends, and acquaintances. One of the differential aspects of an NL and

FL is that most of the knowledge is gained in one's NL through everyday communication. At the same time, FL learning chiefly takes place in schools, and its limited usage restricts the knowledge absorbed. This situation is comparable to the differences between Italian and the regional languages of Venetian or Bergamasque, where regional languages cover fewer topics. NL and Italian are the primary languages for acquiring and exchanging knowledge, while both FL and regional languages are considered subordinate. Therefore, in this study, the decision-making process in Italian is dovetailed to the FL, while decisions in Venetian or Bergamasque dovetail with the FL.

In Study 1, the emotion processing of native Italian-Venetian participants was examined through the web in either Italian or Venetian. Participants rated their emotional intensity from 1 to 7 (1 = very low; 7 = very high) to some phrases and pictures. Since the Venetian is exclusively oral, only spoken phrases or pictures were used. In Task 1, the emotional intensity was rated in response to three kinds of phrases: 1. Reprimands: the phrases parents use to admonish their children, e.g., "Stop that!". 2. Insults, e.g., "You idiot!" and 3. Endearments, e.g., "I love you!". In Task 2, the anchor contraction effect (ACE) was assessed. The ACE is the tendency to report more intense emotions on the emotion rating scale while using the FL (De Langhe, Puntoni, Fernandes & Van Osselaer, 2011). Participants rated the intensity of their specific emotions (e.g., fear) in response to a picture. The results revealed very similar ratings for each emotion in both languages and a lack of language effect. This finding suggests similar emotion processing for these bilinguals in Italian and Venetian languages. The difference between decision-making in these two languages was evaluated in the following two studies.

In Study 2, native Italian-Venetian participants were tested individually by the aural version of the Asian Disease Problem played by computer and the PowerPoint slides of the problem's content. Participants answered the problem either in Italian or Venetian. In the Italian version, the sure option was chosen more when the problem was in the gain frame and chosen less when the problem was in the loss frame. In contrast, this preference was not observed in Venetian. This shows an effect of language: the framing effect was observed in Italian but not in Venetian. In Venetian, there was a stronger preference toward the sure option more frequently, with results in risk aversion reduction.

Study 3 was done by Italian-Venetian bilinguals responding to the Footbridge Dilemma either in Italian or Venetian. The dilemma was presented to the participants in an audio-recording version. Results demonstrated a significant effect of language on the decision-making process in the dilemma; the participants doing the task in Venetian chose a more utilitarian choice of pushing the man off the bridge to save five people, in comparison to the ones using the Italian language.

Study 4 replicated Study 3 with a different group of bilinguals: the native, proficient Italian-Bergamasque speakers. The results of this study replicated those of the previous one; less utilitarian choices were made in Italian than in the local language of Bergamasque. The findings of the last three studies of Miozzoa et al. are in line with the studies evaluating the FLE on the Asian disease problem (e.g., Keysar et al., 2012) and Footbridge Dilemma (Costa et al., 2014; Corey et al., 2017). The findings of this study indicate the extendibility of the FLE observed in the bilinguals

who have one NL and one FL to the balanced bilinguals who have two native languages.

3.6. Circi, Gatti, Russo and Vecchi (2021)

Although various studies have explored the FLE on moral decision-making and risk aversion, the lack of cumulative evidence in this area motivated Circi and colleagues to conduct a meta-analysis on these subjects. They performed a meta-analysis of 47 experiments (38 experiments concentrating on moral decision-making and nine on risk aversion) from 17 studies, which included various native and foreign languages. The FLE has been observed to influence both these subjects in this meta-analysis; regarding moral decision-making, in line with previous studies, an increase in utilitarian choices was shown, and regarding the risk aversion, confirming the previous literature, the FLE on reducing the risk aversion bias was observed.

In addition to the meta-analyses, two meta-regressions were conducted by Circi and colleagues on the moral decision-making studies to evaluate whether the FLE in this domain is moderated by the difference in language proficiency between NL and FL or by the similarity of the two languages. Surprisingly, the proficiency level did not show a moderate effect. This finding contradicts what we have preconceived in this study, and based on the results available in the literature, this lack of effect may seem odd. The authors assume that this finding could have been due to the fact that they were limited in evaluating proficiency at a global level. Hence, they had to take the average proficiency of each sample into account, which reduces the variability of hundreds of participants to a mean value. However, a

moderating effect of NL-FL similarity was observed, and with more similarity between the two languages, less FLE was observed.

3.7. Del Maschio, Crespi, Peressotti, Abutalebi, and Sulpizio (2020)

In a comprehensive meta-analysis, Del Maschio and colleagues (2020) aimed to explore the robustness and magnitude of FLE on decision-making under risk and moral dilemma conditions. Including data from 15 multi-experiment studies, the authors examined the FLE and whether this effect is moderated by language experience (e.g., age of acquisition and proficiency in the FL) or the difference in methodological design features of the experiments (e.g., type of decision problems and task modality).

A reliable FLE was observed in this study, indicating that participants were likelier to make utilitarian choices and unbiased judgments in their FL. However, the authors failed to observe the moderation effect of methodological choices, including problem type, moderating effects of task modality, and personal/impersonal distinction. The moderation effect of language experience on FLE was not observed either. The latter observation aligns with the findings of Circi et al. (2021), in which, surprisingly, the moderating effect of language proficiency on FLE was not observed. However, this finding of Del Maschio and colleagues, as likewise claimed in the work of Circi and colleagues (2021), could be influenced by the fact that they could only evaluate the mean language proficiency of each sample. This evaluation reduces the inter-individual variabilities. In other words, the difference in evaluation methods of language proficiency and the lack of objective proficiency measures

leads to this limitation. As Del Maschio and colleagues recommend, assessing language knowledge and proficiency through objective measures and standardized instruments could lower this limitation, providing more comparable data for the following experimental and meta-analytic studies.

In the following chapter, we will delve into the MouseTracker technique and its features to explain why this method was chosen to explore decision-making in moral dilemmas in the current study.

Chapter 4
The MouseTracker Technique

The subject of decision-making and its cognitive and emotional components have always been an area of interest for psychologists. Revealing the underlying components and decision-making processes can answer many questions about how our mind operates and how we navigate the complexity of making choices and meeting consequences. In the two main previous topics (FLE and moral dilemmas), we discussed how using an FL can affect our decision-making and how moral dilemmas are a context of making a decision. Therefore, decision-making is the common point of these psychological topics of FLE and moral dilemmas. There are various methods, such as eye-tracking and neuroimaging techniques, to study the decision-making process. Here, we discuss why Mouse-tracking tools, specifically the MouseTracker software, among all the other techniques, can be one of the most efficient for exploring this process and why it was chosen for the aim of the current study.

4.1. Mouse-tracking Tools

Traditionally, scientists believed that the cognitive processes and motor systems function independently and that motor movements occur at the end of cognitive processes. The general idea was that the input (perception) initiates the cognitive process, and afterward, the bodily movement occurs as a result. Whereas, now it has been proven that these processes and movements are coextensive. In other

words, the psychical movements are part and parcel with cognitive processes (Freeman, Dale & Farmer, 2011). When a movement decision is made, the neural activity for the selected option gradually increases, and the decision-making process is immediately available to the premotor cortex. Therefore, as the decision unfolds, the neural system guides the movement simultaneously (Cisek & Kalaska, 2005; Cisek & Kalaska, 2010). Hence, gathering data from hand movements while the action takes place can reveal much about internal cognitive processes (Freeman et al., 2011). Mouse-tracking tools measure participants' computer-mouse movements while making choices and provide rich data and a real-time window into the decision-making process. (Stillman, Shen & Ferguson, 2018).

4.2. What Mouse-tracking Tools Measure

Experiments done by Mouse-tracking tools - such as MouseTracker and Mousetrap- usually involve multiple tasks of selecting binary choices. Participants initiate their mouse movement at the bottom center of the computer screen toward one of the two response choices placed at the upper left and upper right corners. They select a response while the computer continuously records their mouse cursor's position at a high sampling rate. The nature of the response options varies depending on the task. However, it typically involves categorizing a presented stimulus into one of two response categories or choosing between two options. This process is repeated across numerous trials, and for each trial, a detailed temporal record is generated, depicting the trajectory of the mouse cursor from the trial's onset until the participant selects a response. The degree of response conflict is achieved by evaluating how directly participants navigate their responses from the starting point to their final

decision. To this aim, the participants' mouse movement trajectory (the actual cursor route) is compared with an imaginary straight path from the beginning to the end of the response (Stillman et al., 2018). The hand movement toward the chosen option provides rich data indicating the temporal development of cognitive processes (Kieslich & Henninger, 2017).

Mouse-tracking techniques offer different metrics to delve into the cognitive processes. One is calculating the area under the curve (AUC), quantifying the response conflict. This phenomenon involves computing the area between the actual cursor trajectory and the hypothetical idealized straight route. The metric known as maximum deviation (MD) also pinpoints the maximum separation between any point on the actual trajectory and the hypothetical straight path. Another assessment is related to the moments when various attributes are melded into mouse movements and evaluating whether trajectories exhibit and unfold sequentially or dynamically. In essence, these metrics examine how choices evolve over time. (Stillman et al., 2018).

4.3. The Benefits of Mouse-tracking Methods in Comparison with Other Techniques

There are several ways other than Mouse-tracking to study the decision-making process, such as reaction times, eye-tracking, functional Magnetic Resonance Imaging (fMRI), and electroencephalography (EEG). All these techniques can offer various reliable data. However, Mouse-tracking provides a rich

variability of data about decision-making and how a decision unfolds while it unfolds with simple materials.

Compared with Mouse Tracker, methods such as eye-tracking, EEG, fMRI, and reaction times can be opaque, expensive, and time-consuming, as well as making complicated attempts to interpret the process of decision unfolding directly. Whereas Mouse-tracking sheds light on the temporal decision unfolding in real-time in a shorter period of time in a user-friendly package. (Stillman et al., 2018).

Another significant benefit of Mouse-tracking is that it can reveal timing details at a millisecond resolution. This feature offers valuable insights into the precise moments when particular factors are processed during a developing decision or the temporal progression of specific processes. In contrast to alternative time-sensitive metrics, Mouse-tracking presents unique advantages. For instance, in choice tasks, eye tracking relies on discrete saccades, which are recorded as rapidly as they occur, typically at a rate of approximately 3-4 times per second, while mouse tracking is based on the continuous movement of the hand in milliseconds (Freeman, 2018). Moreover, outcome-based measures like reaction times (RTs) or error rates also provide data via computer software and are not expensive to use. However, compared to Mouse-tracking, these methods can provide limited interferences of the "sort of perceptual-cognitive processing occurring across time", especially about the process evolvement over time (Freeman & Ambady, 2010, p.226).

4.5. MouseTracker Software

Freeman and Ambady (2010) introduced a version of Mouse-tracking named the MouseTracker technique, based on recording the streaming x- and y-coordinates of the mouse and real-time processing for different cognitive studies. MouseTracker is a cost-free software package that can be installed and used on Windows computers. It can measure real-time hand movements of the mouse trajectories in response to various experimental tasks based on images, letter strings, and/or sounds.

The most important temporal and spatial data this package provides are Response *Conflict* and *The Evolution of Choice*. By gauging the directness of participants' mouse movement toward their choice and comparing it with the hypothetical straight trajectory from the start to the response, the participants' *response conflict* between the two options is revealed; the more similar the participants' trajectories to the straight trajectory, the less conflict between the two options is interpreted. *The Evolution of Choice* is interpreted by quantifying the acceleration and velocity of mouse trajectories toward the answer. This shows whether decision-making unfolds sequentially or dynamically (Stillman et al., 2018).

4.6. How MouseTracker Works

The Mouse Tracker package contains three main programs:

1. **Runner**: The data collection program specifies parameters such as stimuli files, timing, and response options and runs participants through the studies.

2. **Designer**: The graphics-based program can set up the experiments' visual layout and response options based on the researchers' aim.

3. **Analyzer:** The program imports participants' data and visualizes, processes, and analyzes the mouse movements.

Researchers can design graphics-based experiments to determine display, various parameters, lists of stimuli and responses, and timing specifically suited to their research. The design uses a .CSV file in programs such as Excel. To run the experiment, each participant sees a "Start" button on the screen for each trial. The trial begins after pressing the start button. Then, the participant moves the mouse toward the presented options. (Freeman et al., 2011). Depending on what is to be examined, the options vary. However, the two common setups include choosing between two options (e.g., yes or no) or categorizing given stimuli into possible response categories (Stillman et al., 2018). While each experimental trial occurs, the Runner program records between 60 and 75 x-, y-coordinate pairs every second of the mouse movement. It also records *the raw time*, showing how many milliseconds have elapsed. These provide rich data to compute and analyze. After the experiment, the finalized data can be exported as a .CSV file readable in Excel for further analysis (Freeman & Ambady, 2010).

In the next chapter, we will discuss the details of the current study and address how we used this MouseTracker technique to evaluate the effect of the FL on the moral decision-making process.

Chapter 5
The empirical research

In order to discuss the details of the current study, an overview of the study of Lotto, Manfrinat, and Sarlo (2014) is required because the materials we used to explore the FLE on moral dilemmas are the set of dilemmas presented in their work. It is worth mentioning that they did their experiment only in Italian (NL), whereas our study is done using the same dilemmas in Italian (NL) and English (FL).

Lotto et al. (2014). Lotto and colleagues provided a new set of moral dilemmas that dissolves the criticisms of the dilemmas proposed by Greene et al. (2001). One criticism of the dilemmas presented by Greene and colleagues (2001) includes having more emotive language in personal dilemmas than in impersonal ones. Moreover, they included family members or close friends in some dilemmas. These two issues can lead to greater activation in emotional systems, creating emotional confound and bias in responding (Borg, Hynes, Van Horn, Grafton & Sinnott-Armstrong, 2006). Lotto and colleagues solved this problem by applying a plain tone in all dilemmas and avoiding referring to loved ones in their scenarios. Another criticism of the set of dilemmas provided by Greene et al. (2001) is regarding the generalizability of their findings. The item analysis performed by McGuire, Langdon, Coltheart, and Mackenzie (2009) revealed that the significant results Greene et al. (2001) reached were due to a small number of stimuli, hence not generalizable to other populations of moral dilemmas. This issue was addressed in

the work of Lotto et al. (2014) by considering more controlled stimuli and running both subject and item analyses.

Lotto and colleagues presented a new set of 75 dilemmas involving different scenarios. 60 of these dilemmas include deciding whether to kill or cause the death of one person to save the lives of more people, among which 30 are instrumental, and 30 are incidental dilemmas. In instrumental or footbridge-like dilemmas, the death of a person is a means to save the lives of more people. For example, shooting a person or pushing them off the bridge is needed to save other people. Whereas in incidental or trolley-like dilemmas, the death of one person is a predictable but unintended result of the action done to save more people. For example, pushing a button to change the direction of a train from a path toward five people to a path toward one person or guiding a toxic fume from one room where there are five people to a room where there is only one person.

In addition, dilemmas are diverse regarding the Risk-involvement variable in the scenario, including self-involved dilemmas in which the main character's (the respondent's) life is at risk and other-involved dilemmas in which the respondent's life is not at risk. Considering these components, the dilemmas are divided into four conditions: 15 instrumental and self-involved dilemmas, 15 instrumental and other-involved dilemmas, 15 incidental and self-involved dilemmas, and 15 incidental and other-involved dilemmas. They have also added 15 filler dilemmas regarding moral issues that did not include killing anyone and were about moral issues such as stealing and lying. Some of the dilemmas were the redesign of the work of Cushman,

Young, and Hauser (2006), Greene et al. (2001), Greene et al. (2008), and Moore, Clark, and Kane (2008), whereas the rest were newly developed in their study.

The dilemmas had plain language, were balanced for word length, and were presented in Italian in random order to the participants who were Italian native speakers. The participants, 120 undergraduate Italian students, joined the experiment in groups of 10 to 15 students in the same room, while each completed the experiment individually by a computer. They read the dilemmas and their optional responses on the screen and chose their answers between "yes" (as agreeing to do the solution provided) by pushing the right button and "no" (as not agreeing to do the solution provided) by pushing the left button.

They measured the effect of the Type of Dilemma (instrumental or incidental) and Risk-involvement (self-involved or other-involved) on the rate of utilitarian (affirmative) and deontological responses (negative), along with the reaction time to each dilemma. They also assessed the moral acceptability of each dilemma after each response on an 8-point scale and the arousal rating of participants after the experiment on a one to nine-point scale of valence and arousal.

Their results showed more moral acceptance and higher affirmative responses to incidental dilemmas than instrumental ones. Moreover, the decision-making process to engage in the dilemmas, including incidental killing, was slower than the ones including instrumental killing. Regarding the Risk-involvement, the percentage of affirmative responses was higher when the main person's life was at risk. At the same time, the moral acceptancy was lower for these dilemmas.

5.2. The Current Study

In our study, we used Lotto and colleagues' materials using a MouseTracker experiment. We tested two groups of Italian participants; one performed the task in their NL (Italian) and the other group in an FL (English),

5.2.1. Participants

A total of 69 Italian native speakers participated in the experiment, distributed across two groups. Of these, 41 participants (M_age = 24.59, SD_age = 3.50) completed the experiment in Italian, while 28 participants (M_age = 24.46, SD_age = 2.95) completed it in English. Of the 69 participants, Lisa Visentin tested 25 as part of her degree thesis at the University of Padova; Zhimin Hu tested 10, and 34 were tested by myself. The details of the participant pool are shown in Table 5.1

Table 5.1. Descriptive statistics of the participants with Means and Standard deviations in brackets.

	Italian (native condition)	**English (foreign condition)**
Number	41	28
Age	24.59 (3.50)	24.46 (2.95)
Age of acquisition in English	6.15 (1.73)	7.12 (2.28)
Age of fluency in English	16.66 (4.58)	17.39 (3.67)
Age of reading in English	9.56 (3.37)	10.04 (4.31)
Age of reading fluency in English	15.34 (4.32)	16.68 (3.33)
Months spent in an English-speaking country	4.83 (10.95)	2.04 (2.95)
Level of speaking English (out of 10)	7.59 (1.99)	7.14 (1.65)

Level of understanding spoken English (out of 10)	8.41 (1.79)	8.18 (1.47)
Level of understanding written English (out of 10)	9.20 (1.47)	8.82 (1.59)
Self-perceived accentedness in English	4.69 (2.17)	5.14 (1.46)
Other-perceived accentedness in English	6.49 (3.26)	7.50 (2.29)
Exposure to English via friends	5.12 (3.03)	6.00 (2.97)
Exposure to English via family	1.96 (1.82)	1.93 (1.96)
Exposure to English via TV and film	7.00 (3.51)	6.43 (3.17)
Exposure to English via radio and music	8.24 (2.84)	8.71 (1.96)
Exposure to English via reading	8.17 (2.55)	7.50 (2.46)
Exposure to English via self-learning	6.37 (3.55)	7.32 (2.76)

5.2.2. Stimuli and Material

The experiment was designed and run in the MouseTracker software. We applied the set of dilemmas presented by Lotto and colleagues (2014). This set contained 60 dilemmas, including the scenarios of one person's death in favor of saving more lives, including 30 instrumental dilemmas (15 self-involved and 15 other-involve) and 30 incidental dilemmas (15 self-involved and 15 other-involve) and 15 filler dilemmas regarding moral issues other than killing someone. The dilemmas were originally Italian and were translated into English.

The dilemmas were presented written and auditory simultaneously via a computer. Each dilemma and its resolution were displayed in white font on a black background (font: Calibri Light, size: 30). The auditory version was recorded by a synthesized voice in a plain tone and ensured that the participants with different reading speeds spent the same amount of time facing the dilemmas and resolutions.

5.2.3. Procedure

The experiment was done at the University of Padova in the psychology faculty lab. The participants did the test individually on the same computer, with the presence of a research group member to answer the questions and make the procedure straightforward. After receiving the instructions from the research team, participants started the test by reading a slide of a brief instruction. After the instructions, some test scenarios were presented so the participants would familiarize themselves with the task and understand how to make their choices using the mouse.

Each dilemma was shown visually and played auditory, either in English or Italian, by clicking on the start point at the middle-bottom of the screen (see Figure 5.1). The question at the end of each scenario asked the participants whether they would make the choice they had been given. Then, the participants made their choice by moving the mouse and clicking on the "Yes" and "No" options. The "Yes" and “No” options appeared on the top corners of the screen. To control for the influence of the options’ location, the "Yes" and "No" options’ locations were changed after running some experiments. Thus, in 16 English cases and 12 Italian cases, the "Yes" option was placed in the top right corner, and the "No" option was placed in the top

left corner. Whereas the "Yes" option was placed in the top left corner and the "No" option was placed in the top right corner in 13 English and 29 Italian cases.

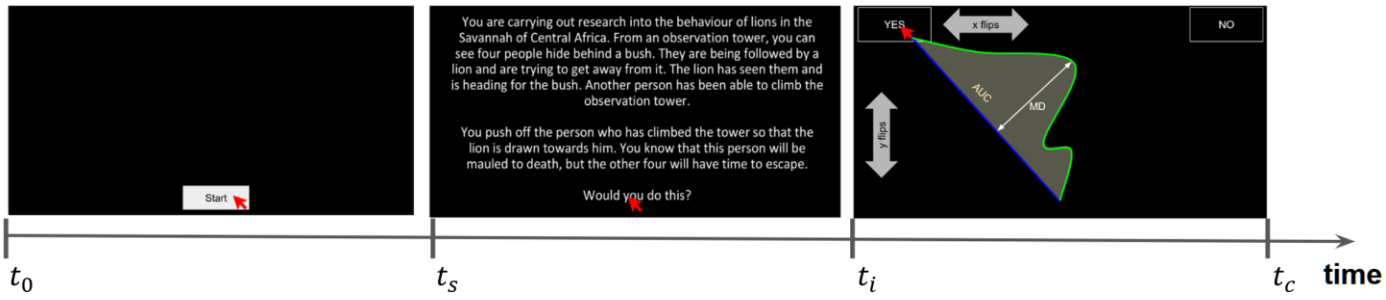

Figure 5.1. The procedure of the test: At t_0, the participants are shown a screen where they can click the start button. When they press the button (at t_s), the screen shows them the dilemma. They are given time to read the problem, and then at t_i, the screen with two options comes up. The participants then click on their answer at time t_c.

Participants could rest and take a break between the dilemmas and did not have a time limitation to answer the test. The mean time of the whole experiment was 50 minutes. At the end of the experiment, participants filled out the Language Experience and Proficiency Questionnaire (LEAP-Q) by Marian, Blumenfeld, & Kaushanskaya (2007) to determine their language proficiency (in both NL and FL), age of acquisition, average spending time on each language, etc. (See Table 5.1.).

5.2.4. Predictions

This research aims to examine the role of language (FL versus NL) in moral judgment, including the type of response and the decisions' trajectories, using the set of dilemmas presented by Lotto et al. (2014). Therefore, we expect to:

1- Replicate the results of Lotto et al. (2014) to observe more utilitarian responses for incidental dilemmas and more deontological ones for instrumental dilemmas in the Italian version.

2 - Replicate the results of Lotto et al. (2014) to observe slower reaction time in response to incidental dilemmas in the Italian version.

3- Observe the FLE on increasing the percentage of utilitarian choices while responding in English.

4- Observe the FLE in the analysis of the trajectories of the decisions in increasing the reaction time and more mouse trajectories while responding in English.

5.3. Results

All the analyses were performed using the R software (R Core Team, 2022) and the mousetrap package (Wulff et al., 2021). A generalized linear mixed model (GLMM) was employed to examine the decision-making processes in moral dilemmas. The GLMM, fit by maximum likelihood (Laplace Approximation), incorporated a binomial family with a logit link function. The model formula accounted for the main effects and interactions of the type of dilemma, risk involvement, and language, alongside random intercepts for subjects (subjID) and dilemmas. The dataset encompassed 4140 observations distributed across 69 unique subjects and 60 unique dilemmas.

The model exhibited an Akaike Information Criterion (AIC) of 3986.2 and a Bayesian Information Criterion (BIC) of 4049.5. The log-likelihood was assessed at -1983.1 with a deviance of 3966.2, and the degrees of freedom for residuals were 4130. The scaled residuals ranged from -6.8175 to 11.4431, suggesting a good model fit to the data.

5.3.1 Decisions

Concerning fixed effects, the type of dilemma surfaced as a significant predictor, with instrumental dilemmas being associated with a decrease in the log odds of the moral response (Estimate = -2.4416, $z = -8.839$, $p < .001$). Additionally, significant interactions were observed between the type of dilemma and language (Estimate = -0.6280, $z = -2.463$, $p = .0138$) and risk involvement and language (Estimate = 0.4434,

z = 2.053, p = .0401), showcasing the moderating effect of language on the relationship between the type of dilemma and risk involvement with the decision-making process. Other interactions and main effects were not statistically significant (see Figure 5.2).

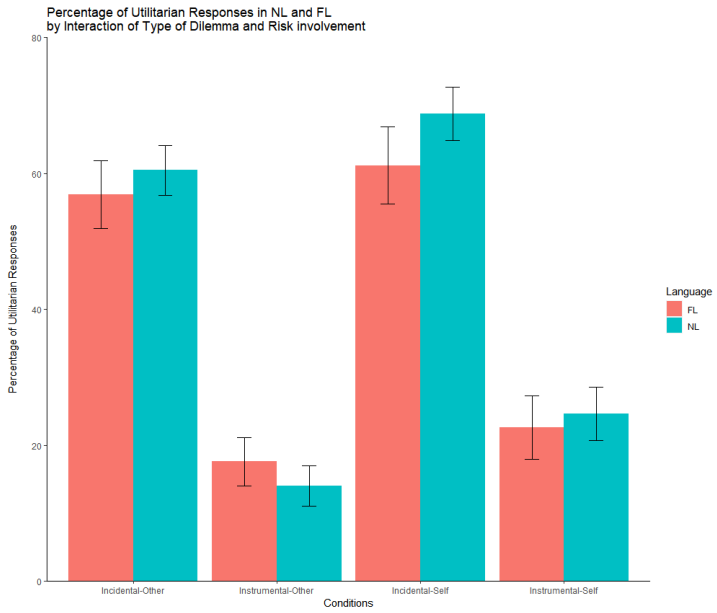

Figure 5.2. Boxplot of the percentage of utilitarian responses in NL and FL

Since the FLE is typically observed in the Footbridge dilemma or, in our case, the instrumental other-involved dilemmas, we performed post hoc analysis to contrast the decisional outcomes in Foreign Language and Native Language under

the four experimental conditions. Within incidental dilemmas, the transition from FL to NL did not significantly affect moral decisions in both Other (Estimate = -0.1719, z = -0.491, p = 0.9997) and Self scenarios (Estimate = -0.6153, z = -1.742, p = 0.6597). Similarly, within instrumental dilemmas, no significant differences were observed between FL and NL, as evidenced in the Other (Estimate = 0.4561, z = 1.215, p = 0.9279) and Self scenarios (Estimate = -0.1896, z = -0.526, p = 0.9995).

5.3.2 Temporal results

Initiation time. In investigating the initiation times in moral dilemmas, a log transformation was performed on the initiation time to attain a more normal distribution, which is conducive to linear modeling. A linear mixed model was then employed to dissect the quadruple interaction of Type of Dilemma, Risk involvement, Language, and Response on the log-transformed initiation times, with random intercepts for subjects and dilemmas. The findings revealed significant main effects of Type of Dilemma (Estimate = -0.59, t = -5.320, $p < 0.001$), indicating shorter initiation times in instrumental dilemmas and Risk involvement (Estimate = -0.26, t = -2.332, p = 0.020), with self-involved dilemmas showing shorter initiation times. A significant interaction was also observed between the Type of Dilemma and Risk involvement (Estimate = 0.41, t = 2.609, p = 0.009), suggesting the influence of dilemma type on initiation time was moderated by the level of self-involvement in the risk. However, the main effect of Language (Estimate = -0.28, t = -1.170, p = 0.242) did not reach significance, nor did its interactions with other predictors, including the Type of Dilemma and Language interaction (Estimate = 0.17, t = 1.221, p = 0.222), the Risk involvement and Language interaction (Estimate = 0.11, t =

0.839, p = 0.401), and the three-way interaction among Type of Dilemma, Risk involvement, and Language (Estimate = -0.19, t = -1.012, p = 0.312). See Table 5.2.

Total decision time. The total decision times in the moral dilemmas were investigated through a log transformation to ensure a normal distribution suitable for linear modeling. A linear mixed model was employed to scrutinize the triple interaction among Type of Dilemma, Risk involvement, and Language on the log-transformed reaction times, accounting for random intercepts for both subjects and dilemmas. The analysis revealed significant main effects for Type of Dilemma (Estimate = -0.37, t = -7.153, p < 0.001) and Risk involvement (Estimate = -0.17, t = -3.350, p = 0.001), indicating shorter reaction times in instrumental dilemmas and self-involved dilemmas respectively. Significant interactions were also observed between the Type of Dilemma and Risk involvement (Estimate = 0.24, t = 3.275, p = 0.001) and between the Type of Dilemma and Language (Estimate = 0.15, t = 2.455, p = 0.014), suggesting a moderated influence of dilemma type on reaction time by the level of risk involvement and language respectively. However, the main effect of Language (Estimate = -0.18, t = -1.630, p = 0.103), the interaction between Risk involvement and Language (Estimate = 0.09, t = 1.469, p = 0.142), and the three-way interaction among Type of Dilemma, Risk involvement, and Language (Estimate = -0.11, t = -1.336, p = 0.182) did not reach significance. The model explained a marginal R^2 of 0.022 and a conditional R^2 of 0.296, indicating a fair amount of variance accounted for by both fixed and random effects. This analysis unveils the factors' complex interplay in modulating reaction times across different dilemma scenarios. See Table 5.2.

Table 5.2. Initiation and Reaction times

Type of dilemma	Risk Involvement	Language	Initiation time (ms) mean (standard deviation)	Reaction time (ms) mean (standard deviation)
Incidental	Other	Foreign	1513 (4500)	3934 (7907)
Instrumental	Other	Foreign	655 (1536)	1944 (2243)
Incidental	Self	Foreign	1136 (3033)	2817 (4854)
Instrumental	Self	Foreign	923 (3590)	2318 (4152)
Incidental	Other	Native	1052 (2745)	2735 (4414)
Instrumental	Other	Native	632 (1819)	2003 (3133)
Incidental	Self	Native	921 (2358)	2426 (4013)
Instrumental	Self	Native	635 (1704)	2001 (2950)

5.3.3 Mouse trajectory results

Maximum deviation (MD). The Maximum Deviation was assessed through a linear mixed model exploring the three-way interaction between Type of Dilemma, Risk involvement, and Language while accounting for random effects associated with the subject and dilemma. The analysis demonstrated significant main effects for the Type of Dilemma (Estimate = -0.07, t = -2.014, p = 0.044) and Risk involvement (Estimate = -0.08, t = -2.177, p = 0.030), suggesting a lesser maximum deviation for instrumental dilemmas than for self-involved scenarios. The main effect of Language did not reach statistical significance (Estimate = -0.08, t = -1.921, p = 0.055, neither did any of the interaction terms: Type of Dilemma × Risk involvement (Estimate = 0.07, t = 1.412, p = 0.158), Type of Dilemma × Language (Estimate = 0.04, t = 0.780, p = 0.436), Risk involvement × Language (Estimate = 0.07, t = 1.501, p = 0.133),

and the three-way interaction among Type of Dilemma, Risk involvement, and Language (Estimate = -0.03, t = -0.532, p = 0.595).

Area under the curve (AUC). The analysis on AUC was conducted using a linear mixed model, evaluating the triple interaction among Type of Dilemma, Risk involvement, and Language, with random effects for subject and dilemma. The main effects showed that Risk involvement (Estimate = -0.08, t = -2.386, p = 0.017) and Language (Estimate = -0.09, t = -2.443, p = 0.015) were significant, denoting a decrease in AUC values for self-involved dilemmas and those presented in NL respectively. The main effect of the Type of Dilemma did not reach statistical significance (Estimate = -0.06, t = -1.848, p = 0.065). The interaction between Type of Dilemma and Risk involvement was significant (Estimate = 0.10, t = 1.972, p = 0.049), indicating a modulation effect between these factors on AUC. However, other interaction terms, including Type of Dilemma × Language (Estimate = 0.04, t = 0.818, p = 0.413), Risk involvement × Language (Estimate = 0.08, t = 1.738, p = 0.082), and the three-way interaction among Type of Dilemma, Risk involvement, and Language (Estimate = -0.06, t = -1.012, p = 0.312) were not statistically significant. See Figure 5.3 for visual representations of each condition's time-normalized average mouse trajectories.

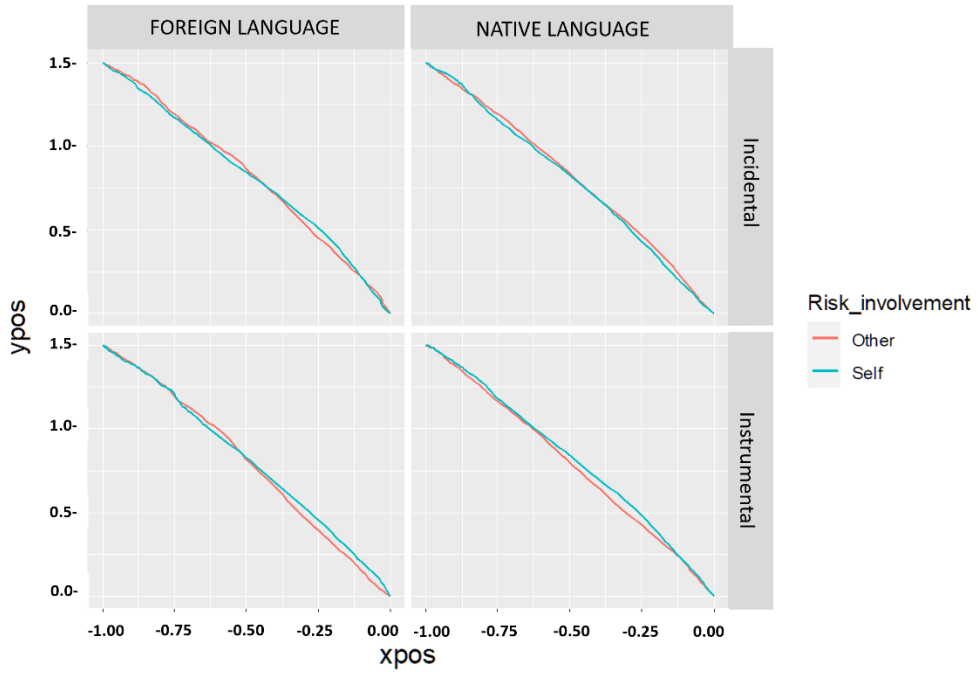

Figure 5.3. Time-normalized average mouse trajectories in different conditions under study.

X flips. The count of directional changes along the x-axis, termed as x flips, was analyzed using a Poisson generalized linear mixed model, considering the triple interaction of Type of Dilemma, Risk involvement, and Language, with random effects for subject and dilemma. The main effects indicated that Type of Dilemma (Estimate = -0.39, z = -5.873, $p < 0.001$), Risk involvement (Estimate = -0.27, z = -4.116, $p < 0.001$), and Language (Estimate = -0.38, z = -3.437, $p < 0.001$) were significant, reflecting a decrease in x flips for instrumental dilemmas, self-involved scenarios, and dilemmas presented in NL. The interaction effects among Type of Dilemma and Risk involvement (Estimate = 0.36, z = 3.867, $p < 0.001$), Type of Dilemma and Language (Estimate = 0.40, z = 5.942, $p < 0.001$), and Risk

involvement and Language (Estimate = 0.26, z = 3.849, p < 0.001) were significant, suggesting that the relationship between these factors modulates x flips. Moreover, the three-way interaction among Type of Dilemma, Risk involvement, and Language was also significant (Estimate = -0.40, z = -4.163, p < 0.001), indicating a complex interplay among these factors in influencing x flips. This analysis elucidates how the multifaceted interaction of dilemma type, risk involvement, and language can significantly impact the decision-making trajectory as indicated by x flips, thereby providing a nuanced understanding of the cognitive processing involved in moral judgments.

Y flips. The count of directional changes along the y-axis, termed y flips, was analyzed using a Poisson generalized linear mixed model, considering the triple interaction of Type of Dilemma, Risk involvement, and Language, with random effects for subject and dilemma. The main effects demonstrated that Type of Dilemma (Estimate = -0.40, z = -5.425, p < 0.001), Risk involvement (Estimate = -0.35, z = -4.783, p < 0.001), and Language (Estimate = -0.38, z = -3.030, p = 0.002) were significant, indicating a decrease in y flips for instrumental dilemmas, self-involved scenarios, and dilemmas presented in NL. The interaction effects among Type of Dilemma and Risk involvement (Estimate = 0.38, z = 3.562, p < 0.001), Type of Dilemma and Language (Estimate = 0.35, z = 4.795, p < 0.001), and Risk involvement and Language (Estimate = 0.22, z = 3.022, p = 0.003) were significant, suggesting that the relationship between these factors modulates y flips. Moreover, the three-way interaction among Type of Dilemma, Risk involvement, and Language

was also significant (Estimate = -0.26, z = -2.481, p = 0.013), indicating a complex interplay among these factors in influencing y flips.

Maximum velocity. The analysis of maximum velocity was approached using a Gamma generalized linear mixed model, considering the triple interaction of Type of Dilemma, Risk involvement, and Language, with random effects for subject and dilemma. The model revealed that the main effects of Type of Dilemma (Estimate = -0.018, t = -0.619, p = 0.536), Risk involvement (Estimate = -0.031, t = -1.045, p = 0.296), and Language (Estimate = -0.109, t = -1.356, p = 0.175) were not statistically significant. Similarly, none of the interaction terms were significant: Type of Dilemma × Risk involvement (Estimate = 0.011, t = 0.262, p = 0.794), Type of Dilemma × Language (Estimate = 0.024, t = 0.753, p = 0.451), Risk involvement × Language (Estimate = 0.001, t = 0.046, p = 0.963), and the three-way interaction among Type of Dilemma, Risk involvement, and Language (Estimate = 0.012, t = 0.277, p = 0.782). This suggests that, unlike previous measures, the maximum velocity data did not exhibit a significant modulation by these factors.

Time to reach maximum velocity. The analysis for the time to reach maximum velocity was conducted using a linear mixed model, considering the triple interaction of Type of Dilemma, Risk involvement, and Language, alongside random effects for subject and dilemma. The model revealed significant main effects of Type of Dilemma (Estimate = -1350.92, t = -5.230, $p < 0.001$), Risk involvement (Estimate = -620.84, t = -2.403, p = 0.016), and Language (Estimate = -773.74, t = -2.065, p = 0.039), indicating that these factors significantly affected the time it took to reach

maximum velocity. The negative estimates suggest that the time to reach maximum velocity is lower for instrumental dilemmas, self-related risk involvement, and the non-NL language condition compared to their respective counterparts. Furthermore, significant interactions were found between the Type of Dilemma and Risk involvement (Estimate = 947.43, t = 2.593, p = 0.010) and between the Type of Dilemma and Language (Estimate = 757.70, t = 2.403, p = 0.016). These interactions suggest a complex relationship among the factors influencing the time to reach maximum velocity, with higher values observed when instrumental dilemmas are combined with self-related risk involvement and when instrumental dilemmas are combined with the FL condition. The three-way interaction among Type of Dilemma, Risk involvement, and Language was not significant (Estimate = -702.97, t = -1.577, p = 0.115).

Chapter 6
Discussion of the Results

Various studies have shown the effect of an FL on decision-making, including moral judgments. However, some aspects and details of this decision-making process have remained unrevealed, such as the different temporal features and the interaction between the FL and other variables in the decision-making process. In this study, we have examined the effect of the FL on native Italian speakers on different aspects of moral decision-making. Using the MouseTracker software, the set of dilemmas developed by Lotto et al. (2014) was presented to the participants in Italian (NL) or English (FL). MouseTracker records and analyses the movements of the computer mouse while the decision is being made; therefore, it offers a continuous ongoing stream of rich cognitive outputs and reveals the time course of cognitive processes (Freeman & Ambady, 2010). We have measured the effects of the Language (NL or FL), Risk Involvement (self-involved or other-involved), and the Type of Dilemma (instrumental or incidental), as well as their two-way and three-way interactions on moral decision-making, along with various temporal aspects.

As predicted, our results replicated the findings of Lotto et al. (2014), with a decrease in utilitarian choices in instrumental dilemmas and an increase in utilitarian choices in incidental dilemmas. This effect was present in both FL and NL. On the other hand, despite the findings of Costa et al., 2014, Hayakawa et al., 2017, and Miozzoa et al., 2020, we failed to find a main effect of FL on increasing the utilitarian

responses. Based on our findings, FL remains a moderator for the impact of the Type of Dilemmas and Risk Involvement on decision-making, in the sense that the difference in utilitarian responses for participants doing the test in FL is less than the ones doing the test in their NL when changing the Type of Dilemma to instrumental. This finding aligns with previous literature focusing on the interaction of FL and the Type of Dilemma (Cipolletti et al., 2016; Corey et al., 2017). However, our results did not replicate the significant increase in utilitarian choices in response to instrumental dilemmas while using the FL. Moreover, this difference in change was also found when the dilemma puts others' lives at risk for the participants doing the test in FL.

Regarding the temporal results, the *initiation time* and *total decision time* were assessed. We observed shorter initiation time as a function of the Type of Dilemma and Risk Involvement. This shorter initiation time was observed in instrumental dilemmas and self-involved dilemmas, indicating that participants rely on the first system activation (Stillman, 2018), an automatic, emotional, intuitive, and unconscious system (Kahneman, 2003) in response to these dilemmas. However, an effect of language on initiation time was not observed directly, nor in interaction with the Type of Dilemma and Risk Involvement. As to the total decision time, shorter reaction times in instrumental dilemmas and self-involved dilemmas were observed. The Type of Dilemma also moderated, reducing the reaction time for Risk Involvement and Language. However, the Language did not affect the reaction time.

Maximum deviation (MD), Area under the curve (AUC), X flips, Y flips, Maximum velocity, and time to reach maximum velocity (MVT) were assessed for Mouse trajectory. *Maximum deviation (MD)* refers to the maximum distance between the peak of the actual mouse trajectory and a straight hypothetical trajectory from the start point to the made choice. The *area under the curve (AUC)* is the area between the actual trajectory and the straight hypothetical trajectory. These two metrics are indicators of conflict in decision-making (Stillman et al., 2018).

Among these metrics, MD shows little to no correlation regarding the language effect, and our results, based on the AUC values, suggest less conflict in decision-making when participants are presented with either self-involved, instrumental, or dilemmas in their NL.

X-flips and Y-flips are the numbers of directional changes along the X and Y axes, showing the complexity of the decision-making process (Freeman & Ambady, 2010) and uncertainty in the response (Koop, 2013). More X-flips and Y-flips in incidental dilemmas, other-involved dilemmas, and dilemmas presented in the FL show that participants had more conflict and complexity in the said dilemmas and in a non-native language. Our results contrast the work of Koop (2013), who failed to find any relationship between the Type of Dilemma and x-flips.

Regarding the velocity, we failed to replicate the findings of Parker and Finkbeiner (2020), who found a greater maximum velocity for utilitarian responses in personal (instrumental) dilemmas, indicating that participants were more hesitant in making

such a decision. However, we observed a greater MVT in incidental, other-involved, and dilemmas presented in the FL. These findings suggest that the participants needed more time to make utilitarian decisions in the abovementioned dilemmas.

Conclusion

Several studies have applied various methods to explore the ELE topic in different psychological areas. Although the robustness of this effect has generally been proved, some aspects and details of this effect are still controversial or have yet to be revealed. In this study, utilizing the MouseTracker method, we have delved into the procedure of FLE on moral decision-making and the potential influential factors through which the FLE emerges. MouseTracker records and analyzes mouse movements while the decision is being made and chosen, providing rich cognitive outputs and revealing the real-time cognitive process (Freeman & Ambady, 2010). These features make this method a valuable technique to investigate the decision-making processes.

Designing and applying the experiment by MouseTracker software, we presented the dilemmas developed by Lotto et al. (2014) to 69 native Italian speakers. A group of 41 participants did the task in Italian (NL), and a group of 28 did it in English (FL). This set of dilemmas includes scenarios that are diverse regarding the Risk Involvement (self-involved: when the respondent's life is at risk/other-involved when the lives of other people are at risk) and Type of the Dilemma (instrumental: when the death of a person is a means to save more lives/ incidental: when the death of a person is an unintended but predictable result of saving more lives). Presenting these dilemmas in two languages allowed us to explore the FLE on moral decision-

making by evaluating the effects of the Language, Risk Involvement, and the Type of Dilemma, as well as their two-way and three-way interactions.

Even though we could not replicate the general assumption and observation of the direct effect of FL on increasing the utilitarian responses in instrumental (footbridge-like) dilemmas in the literature, by adopting the perspectives mentioned above, we managed to highlight the complex and multi-dimensional role of language in decision-making and, precisely, in the moral judgment process. We could also underline the impact of FL on increasing hesitance and complexity while decision-making.

To our knowledge, the present study is the first to investigate the role of Risk Involvement in bilinguals' moral decision-making. Risk Involvement is an influential confounding variable, and its role should be taken into account. Although the role of friends and family members in the previous dilemmas has been considered confounding, the role of self-involvement has yet to be explored in more detail. There is an inclination toward utilitarian choices when the lives of familiar people are at risk in a dilemma. Therefore, it can be predictable and expected for this inclination to be more significant when the scenario involves the respondent's life at risk.

Our study also underscores the features and advantages of utilizing the MouseTracker technique in investigating different aspects of moral decision-making and the FLE. MouseTracker allows us to explore the role of various influential

variables precisely and comprehensively, delving into the complexity of cognitive processes.

Acknowledgment

I sincerely thank Professor Eduardo Navarrete for all the guidance and support I received from him in every step of writing this thesis and Zhimin Hu for his help and support. I also thank my partner, friends, and family for supporting me through this project.

References

1. Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, *5*(SEP). https://doi.org/10.3389/fpsyg.2014.01055
2. Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The Moral Foreign-Language Effect. *Philosophical Psychology*, *29*(1), 23–40. https://doi.org/10.1080/09515089.2014.993063
3. Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, *28*(4), 1131–1141. https://doi.org/10.3758/s13423-020-01871-z
4. Cisek, P., & Kalaska, J. F. (2005). Neural Correlates of Reaching Decisions in Dorsal Premotor Cortex: Specification of Multiple Direction Choices and Final Selection of Action. *Neuron*, *45*(5), 801–814. https://doi.org/10.1016/j.neuron.2005.01.027
5. Cisek, P., & Kalaska, J. F. (2010). Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience*, *33*(1), 269–298. https://doi.org/10.1146/annurev.neuro.051508.135409

6. *Core, R. (2022). Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online: https://www. R-project. org.* (n.d.).
7. 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
8. 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
9. Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment. *Psychological Science*, *17*(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x
10. De Langhe, B., Puntoni, S., Fernandes, D., & van Osselaer, S. M. J. (2011). The Anchor Contraction Effect in International Marketing Research. *Journal of Marketing Research*, *48*(2), 366–380. https://doi.org/10.1509/jmkr.48.2.366
11. del Maschio, N., Crespi, F., Peressotti, F., Abutalebi, J., & Sulpizio, S. (2022). Decision-making depends on language: A meta-analysis of the Foreign Language Effect. *Bilingualism: Language and Cognition*, *25*(4), 617–630. https://doi.org/10.1017/S1366728921001012
12. 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. https://doi.org/10.1080/01434630408666529

13. Dewaele, J.-M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics*, *40*(10), 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002
14. Díaz-Lago, M., & Matute, H. (2019). Thinking in a Foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, *72*(1), 41–51. https://doi.org/10.1177/1747021818755326
15. Foot, P. (2002). The Problem of Abortion and the Doctrine of the Double Effect. In *Virtues and Vices* (pp. 19–32). Oxford University Press. https://doi.org/10.1093/0199252866.003.0002
16. Freeman, J. B. (2018). Doing Psychological Science by Hand. *Current Directions in Psychological Science*, *27*(5), 315–323. https://doi.org/10.1177/0963721417746793
17. Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226
18. Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00059

19. Gao, C. (2013). A Sociolinguistic Study of English Taboo Language. *Theory and Practice in Language Studies*, *3*(12). https://doi.org/10.4304/tpls.3.12.2310-2314
20. Gao, S., Zika, O., Rogers, R. D., & Thierry, G. (2015). Second Language Feedback Abolishes the "Hot Hand" Effect during Even-Probability Gambling. *The Journal of Neuroscience*, *35*(15), 5983–5989. https://doi.org/10.1523/JNEUROSCI.3622-14.2015
21. Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, *59*. https://doi.org/10.1016/j.jesp.2015.02.001
22. Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6
23. 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
24. 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*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004
25. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral

Judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

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

27. Hakuta, K. (2009). Bilingualism. In *Encyclopedia of Neuroscience* (pp. 173–178). Elsevier. https://doi.org/10.1016/B978-008045046-9.01877-5

28. Harris, C. L., Ayçíçeğí, 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

29. Harris, J. G., Cullum, C. M., & Puente, A. E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults. *Journal of the International Neuropsychological Society*, *1*(1), 10–16. https://doi.org/10.1017/S1355617700000059

30. 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

31. Incera, S., Tuft, S. E., Fernandes, R. B., & McLennan, C. T. (2020). Using mouse tracking to investigate auditory taboo effects in first and second

language speakers of American English. *Cognition and Emotion*, *34*(6). https://doi.org/10.1080/02699931.2020.1741339

32. Ivaz, L., Costa, A., & Duñabeitia, J. A. (2016). The emotional impact of being myself: Emotions and foreign-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 489–496. https://doi.org/10.1037/xlm0000179

33. Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

34. Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

35. Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The Foreign-Language Effect. *Psychological Science*, *23*(6), 661–668. https://doi.org/10.1177/0956797611432178

36. Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

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

38. Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*(5), 527–539. https://doi.org/10.1017/S1930297500003636
39. Kvalnes, Ø. (2019). *Moral Reasoning at Work*. Springer International Publishing. https://doi.org/10.1007/978-3-030-15191-1
40. Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A New Set of Moral Dilemmas: Norms for Moral Acceptability, Decision Times, and Emotional Salience. *Journal of Behavioral Decision Making*, *27*(1), 57–65. https://doi.org/10.1002/bdm.1782
41. Marian, Viorica, Henrike K. Blumenfeld, and Margarita Kaushanskaya. "The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals." (2007).McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, *45*(3), 577–580. https://doi.org/10.1016/j.jesp.2009.01.002
42. Miozzo, M., Navarrete, E., Ongis, M., Mello, E., Girotto, V., & Peressotti, F. (2020). Foreign language effect in decision-making: How foreign is it? *Cognition*, *199*, 104245. https://doi.org/10.1016/j.cognition.2020.104245
43. Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who Shalt Not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment. *Psychological Science*, *19*(6), 549–557. https://doi.org/10.1111/j.1467-9280.2008.02122.x

44. Oganian, Y., Korn, C. W., & Heekeren, H. R. (2016). Language switching—but not foreign language use per se—reduces the framing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 140–148. https://doi.org/10.1037/xlm0000161

45. Parker, S., & Finkbeiner, M. (2020). Examining the unfolding of moral decisions across time using the reach-to-touch paradigm. *Thinking & Reasoning*, *26*(2), 218–253. https://doi.org/10.1080/13546783.2019.1601640

46. Rosselli, M., Ardila, A., Lalwani, L. N., & Vélez-Uribe, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *19*(3), 489–503. https://doi.org/10.1017/S1366728915000309

47. Schaich Borg, J., Hynes, C., van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation. *Journal of Cognitive Neuroscience*, *18*(5), 803–817. https://doi.org/10.1162/jocn.2006.18.5.803

48. Spivey, M. J., & Dale, R. (2006). Continuous Dynamics in Real-Time Cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

49. Sulpizio, S., Toti, M., del Maschio, N., Costa, A., Fedeli, D., Job, R., & Abutalebi, J. (2019). Are you really cursing? Neural processing of taboo words in native and foreign language. *Brain and Language*, *194*, 84–92. https://doi.org/10.1016/j.bandl.2019.05.003

50. Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish–English bilingual speakers. *Cognition and Emotion*, *21*(5), 1077–1090. https://doi.org/10.1080/02699930601054133
51. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7
52. Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*, *94*(6), 1395. https://doi.org/10.2307/796133
53. 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
54. Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (n.d.). *Movement tracking of psychological processes: A tutorial using mousetrap*. https://pascalkieslich.github.io/mousetrap-resources/

Supplementary Materials

A1. Language Experience and Proficiency Questionnaire (LEAP-Q)

Language:

This is my [native/first/second/third/fourth/fifth] language.

All questions below refer to your knowledge of [language]:

(1) Age when you…:

began acquiring [language]:	*became fluent* in [language] :	*began reading* in [language] :	*became fluent reading* in [language] :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where [language] is spoken		
A family where[language] is spoken		
A school and/or working environment where [language] is spoken		

(3) On a scale from zero to ten, please select your *level of* ***proficiency*** in speaking, understanding, and reading [language] from the scroll-down menus:

Speaking		Understanding spoken language		Reading	

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning [language] :

Interacting with friends		Language tapes/self instruction	
Interacting with family		Watching TV	
Reading		Listening to the radio	

(5) Please rate to what extent you are currently exposed to [language] in the following contexts:

Interacting with friends		Listening to radio/music	
Interacting with family		Reading	
Watching TV		Language-lab/self-instruction	

(6) In your perception, how much of a foreign accent do you have in [language] ?

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in [language] :

This is my [native/first/second/third/fourth/fifth] language.

All questions below refer to your knowledge of [language]:

(1) Age when you…:

began acquiring [language]:	*became fluent* in [language] :	*began reading* in [language] :	*became fluent reading* in [language] :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where [language] is spoken		
A family where[language] is spoken		
A school and/or working environment where [language] is spoken		

(3) On a scale from zero to ten, please select your *level of* ***proficiency*** in speaking, understanding, and reading [language] from the scroll-down menus:

Speaking		Understanding spoken language		Reading	

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning [language] :

Interacting with friends		Language tapes/self instruction	
Interacting with family		Watching TV	
Reading		Listening to the radio	

(5) Please rate to what extent you are currently exposed to [language] in the following contexts:

Interacting with friends		Listening to radio/music	
Interacting with family		Reading	
Watching TV		Language-lab/self-instruction	

(6) In your perception, how much of a foreign accent do you have in [language] ?

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in [language] :