

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/312177389>

Our Moral Choices Are Foreign to Us

Article in *Journal of Experimental Psychology Learning Memory and Cognition* · January 2017

DOI: 10.1037/xlm0000356

CITATION

1

READS

274

7 authors, including:



Joanna D. Corey

University Pompeu Fabra

3 PUBLICATIONS 3 CITATIONS

SEE PROFILE



Alice Foucart

University Pompeu Fabra

23 PUBLICATIONS 401 CITATIONS

SEE PROFILE



Melina Aparici

Autonomous University of Barcelona

27 PUBLICATIONS 228 CITATIONS

SEE PROFILE



Albert Costa

University Pompeu Fabra

208 PUBLICATIONS 6,718 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Syllable processing and syllabic representation [View project](#)



Typewriting [View project](#)

Journal of Experimental Psychology: Learning, Memory, and Cognition

Our Moral Choices Are Foreign to Us

Joanna D. Corey, Sayuri Hayakawa, Alice Foucart, Melina Aparici, Juan Botella, Albert Costa, and Boaz Keysar

Online First Publication, January 9, 2017. <http://dx.doi.org/10.1037/xlm0000356>

CITATION

Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017, January 9). Our Moral Choices Are Foreign to Us. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0000356>

Our Moral Choices Are Foreign to Us

Joanna D. Corey
Universitat Pompeu Fabra

Sayuri Hayakawa
University of Chicago

Alice Foucart
Universitat Pompeu Fabra

Melina Aparici
Universitat Autònoma de Barcelona

Juan Botella
Universidad Autónoma de Madrid

Albert Costa
Universitat Pompeu Fabra and Institutió Catalana de Recerca i
Estudis Avançats

Boaz Keysar
University of Chicago

Though moral intuitions and choices seem fundamental to our core being, there is surprising new evidence that people resolve moral dilemmas differently when they consider them in a foreign language (Cipolletti et al., 2016; Costa et al., 2014a; Geipel et al., 2015): People are more willing to sacrifice 1 person to save 5 when they use a foreign language compared with when they use their native tongue. Our findings show that the phenomenon is robust across various contexts and that multiple factors affect it, such as the severity of the negative consequences associated with saving the larger group. This has also allowed us to better describe the phenomenon and investigate potential explanations. Together, our results suggest that the foreign language effect is most likely attributable to an increase in psychological distance and a reduction in emotional response.

Keywords: foreign language, bilingualism, moral psychology, decision making

In March of 1841, 41 survivors of a sinking ship occupied a leaky lifeboat as a raging storm intensified around them. It became clear that the overloaded boat would sink with all aboard, so the crew threw the adult males into the sea to save the women and children. When the remaining survivors were rescued, one of the sailors, Alexander Holmes, was charged with murder (*United States v. Holmes*).

How we respond to these moral dilemmas is thought to reflect our core moral fabric, deeply held values, and convictions of right

and wrong. There is new evidence suggesting that these responses can be influenced by the language context. Specifically, when using a foreign language (FL) compared with a native language (NL), people are much more likely to choose to sacrifice one person to save many people (Cipolletti, McFarlane, & Weissglass, 2016; Costa et al., 2014a; Geipel, Hadjichristidis, & Surian, 2015).

We conducted several studies aimed at assessing (a) the robustness of the foreign language effect (FLE) in several contexts and (b) the influence of various factors on the presence of the FLE.

Joanna D. Corey, Center for Brain and Cognition, Universitat Pompeu Fabra; Sayuri Hayakawa, Department of Psychology, University of Chicago; Alice Foucart, Center for Brain and Cognition, Universitat Pompeu Fabra; Melina Aparici, Departament de Psicologia Bàsica, Evolutiva i de l'Educació, Universitat Autònoma de Barcelona; Juan Botella, Facultat de Psicologia, Universidad Autónoma de Madrid; Albert Costa, Center for Brain and Cognition, Universitat Pompeu Fabra, and Institutió Catalana de Recerca i Estudis Avançats; Boaz Keysar, Department of Psychology, University of Chicago.

All authors contributed to study concept and design. Data collection was performed by Joanna D. Corey, Melina Aparici, and Alice Foucart. Data analysis was completed by Joanna D. Corey, Sayuri Hayakawa, and Alice Foucart under the supervision of Juan Botella. All authors contributed to drafting the manuscript and approved the final version for submission.

This research was partially funded by grants from the Spanish Government (PSI2011-23033, Consolider Ingenio 2010 CSD2007-00048, and

Ministry of Economy and Competitiveness PSI2014-52181-P), from the Catalan Government (SGR 2009-1521), from the European Research Council under the European Community's Seventh Framework (FP7/2007-2013 Cooperation grant agreement 613465-AThEME), and from the John Templeton Foundation and the National Science Foundation #1520074 to the University of Chicago. Joanna D. Corey was supported by a grant from the Catalan Government (FI-DGR).

The authors declare that they have no conflicts of interest with respect to their authorship or the publication of this article.

We thank the professors who helped us collect the data by giving us access to their students. We would also like to thank our colleagues for their comments, and Leigh Burnett and Candice Frances for editorial suggestions. All materials are available upon request from the corresponding author.

Correspondence concerning this article should be addressed to Joanna D. Corey, Center for Brain and Cognition, Universitat Pompeu Fabra, c/Tànger 122, Office 55.128, 08018 Barcelona, Spain. E-mail: joannacorey@gmail.com

These studies help to provide a more complete understanding of the phenomenon and evaluate potential explanations for it.

Moral Dilemmas in a NL and a FL

Most of the studies assessing people's choices in moral dilemmas have been conducted in the participants' NL.¹ In such studies, people show a strong aversion to opt for a choice in which one person's life is sacrificed to save many people, such as the footbridge dilemma (Thomson, 1985), a variant of the trolley problem or switch dilemma (Foot, 1978). In this dilemma, you imagine you are standing on a footbridge above a train track. An out-of-control trolley is heading toward five people. The only way to stop the trolley from killing these five people is to push a heavy man off the bridge and in front of the trolley. You have to decide if you would push the man to his death to save the five people. The great majority of people refuse to do this on the grounds that it violates their core morality (Cushman, Young, & Hauser, 2006; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Thomson, 1986; Valdesolo & DeSteno, 2006). Thus, in the footbridge dilemma, most people choose not to sacrifice one to save five.

However, people's choices change with relatively small variations in the context. For instance, in the switch dilemma, you can save the five people by pulling a lever to switch the train to a different track. There is one man on the other track, so switching the track will kill him. Although the consequences of action in the footbridge and switch dilemmas are the same, killing one would save five, the great majority of people choose this option in switch but not footbridge (e.g., Greene et al., 2001).

Much research has investigated why people's choices are so different in these two cases as well as in conceptually similar ones (e.g., Moore, Clark, & Kane, 2008; Nakamura, 2013; Nichols & Mallon, 2006). For instance, people are less likely to sacrifice the one man when this involves exerting personal force, for example, pushing him, relative to when it does not, for example, pushing a button, flipping a switch, or using a pole to push the man (Greene et al., 2009). People also find this option less permissible when the man's body stops the train so that killing the man is instrumental (the direct means) to saving the five people, as opposed to the man's death being a side effect of switching the track (e.g., Cushman, Young, & Hauser, 2006). According to Lieberman (2007), the neuroimaging results from Greene et al. (2001) are consistent with the notion that people weigh their personal involvement in the decision-making process more heavily in contexts like the footbridge dilemma than those like switch. Indeed, the dilemmas vary in the potential repercussions of taking action and, consequently, they may be processed differently. In the footbridge dilemma, choosing the utilitarian option makes one guilty of intentional homicide, whereas in the switch dilemma this is not immediately obvious and therefore may be easier to justify—switching the track saves five lives and only afterward does the person on the other track die, a detail about which one could feign ignorance.² Finally, one of the key differences between these two dilemmas may be the intuitiveness of choosing to sacrifice one to save five. That is, in the switch dilemma, the choice to sacrifice one person is intuitive in that it is instantly compelling to most, whereas the same choice is counterintuitive in the footbridge dilemma due to emotional conflict (Kahane et al., 2012). Indeed,

people tend to make the intuitive choice in both dilemmas—utilitarian for switch and deontological for footbridge—which is consistent with the assertion that these decisions are largely automatic or intuitive (e.g., Haidt, 2001).

However, there is another factor that seems to modulate the likelihood of choosing to sacrifice one to save five. Importantly, this factor is orthogonal to the content: the language in which the dilemma is presented. Costa et al. (2014a) showed that willingness to push the man to his death more than doubles when people consider the dilemma in a FL compared with a NL. They demonstrated this effect using English, Hebrew, and Korean as a NL, and English, French, and Spanish as a FL. In addition, the effect was present for native English speakers using Spanish and for native Spanish speakers using English, showing that indeed the foreignness of the language is what affected people's choices. Cicolletti et al. (2016) and Geipel et al. (2015) replicated the basic effect with several language populations and found a remarkably similar pattern: Choices to save the larger number of people doubled when using a FL relative to the NL for footbridge but not switch. Thus, the existing studies demonstrate that the effect generalizes across different language pairings, and as such it cannot be explained by cultural norms.

Potential Explanations for the Effect of Language

Although the FLe has been well-established, there is scarce evidence regarding its origins or how it is modulated by factors that affect people's choices in NL contexts. In what follows, we review various factors that lead to an increase in utilitarian choices in general and then report differences between NL and FL processing that could account for the presence of the FLe. We then outline various explanations that may account for the effect and introduce the studies we conducted to assess them.

Typically, the choice to sacrifice few lives to save many has been described as "utilitarian," as it is consistent with a preference for maximizing overall welfare, whereas the choice to not do so has been described as "deontological" because this is consistent with a preference for adhering to moral rules such as "do not kill." However, such choices need not necessarily be indicative of the motivation that drives them.

Choosing to sacrifice one person to save many may indicate actual utilitarian inclination of an impartial concern for the greater good of humanity, but it may also indicate a variety of other things (see Kahane, 2015; Kahane, Everett, Earp, Farias, & Savulescu, 2015). For example, such choices are associated with traits that are

¹ It is well known that the overwhelming majority of research participants in the field of psychology are undergraduate students, most often from the United States (Arnett, 2008). Given that international students make up only 4% of all university students in the United States (Open Doors, 2014), it can be assumed that most research is conducted in the participants' native tongue.

² There are still further differences between the dilemmas. Such as whether the action required to save five involves a morally relevant object (person) or not (switch), the locus of intervention is either a morally relevant object (person) or not (train) (Waldmann & Dieterich, 2007), the victim is described as a person or compared with a heavy object, and whether the action would be acceptable in another context (switching a train track) or not (pushing a person; see Waldmann, Nagel, & Wiegmann, 2012). These differences may contribute to the intuitiveness of the utilitarian choice varying across the dilemmas.

consistent with enhanced cognitive control such as increased need for cognition (Conway & Gawronski, 2013; Wiech et al., 2013) and increased working memory capacity (Moore et al., 2008). Enhanced cognitive control would prompt choosing the option best suited by the cost–benefit analyses, which would be congruent with the simple explanation that it is better to save a greater number of lives (Kahane et al., 2012, 2015). Alternatively, choosing to sacrifice one person could indicate a reduction in social or emotional processing such as reduced aversion to causing harm (Bartels & Pizarro, 2011; Cushman, Gray, Gaffey, & Mendes, 2012; Wiech et al., 2013) or reduced empathy (Choe & Min, 2011). Furthermore, deontological and utilitarian inclinations are affected by different factors. Thus, although utilitarian choices are associated with cognitive control, deontological choices are associated with empathic concern, perspective taking, and religiosity (Conway & Gawronski, 2013). Therefore, it could also be that a so-called utilitarian choice may not indicate strong utilitarian inclinations, but instead weak deontological ones (Conway & Gawronski, 2013). This account is consistent with Greene's (2007) dual process model in that the two systems in question are separate, affected by different factors, and that which is most active will determine the response.

Thus, our focus is on the willingness to sacrifice one to save five, independent of whether it is due to decreased deontology or increased utilitarianism. Both accounts of moral choice would predict that the use of a FL would decrease the likelihood of responding in a manner consistent with automatic processing (i.e., emotional [Greene et al., 2001] or intuitive [Kahane et al., 2012]) and increase the likelihood of responding in a manner consistent with controlled processing. Furthermore, both accounts would predict an effect of language only for decisions that elicit strong emotional conflict, which would make it such that choosing to sacrifice one to save five is not the intuitive response and, therefore, this choice may require more controlled processing. For example, in footbridge the utilitarian choice requires the use of personal force and the person's death is instrumental, but this is not the case for the switch dilemma. These two differences between the dilemmas arguably contribute to stronger emotional conflict and the utilitarian response being less intuitive in the footbridge dilemma. Furthermore, these differences between the dilemmas have been argued to be critical in eliciting different moral intuitions and response tendencies (e.g., Cushman, 2013; Railton, 2016). First, people may have learned through experience that the use of personal force leads to bad outcomes, and therefore considering applying such force elicits a negative response. Second, instrumental harm is typically considered to be worse than equivalent harm that occurs as a side effect (e.g., Foot, 1978; Cushman et al., 2006), and individuals who find the former acceptable (representing harm as a subordinate goal) are often considered to be deficient in emotion-based processing (e.g., Koenigs et al., 2007). This would explain why the phenomenon has been consistently found in the footbridge dilemma but not in the switch dilemma. We explain this in more detail in the following paragraphs.

The idea is that using a FL, as opposed to a NL, would prompt more controlled processing, a reduced emotional response, and an increase in psychological distance. Independent evidence suggests that all of these factors can lead to an increase in utilitarian choices in dilemmas that involve substantial emotional conflict. Such

choices have been associated with controlled processing (Greene et al., 2001; Greene et al., 2004; Kahane et al., 2012), with compromised emotional processing (e.g., in individuals with brain lesions in areas that regulate emotion, see Koenigs et al., 2007; in healthy intoxicated individuals, see Duke & Bègue, 2015) and with psychological distance (Aguilar, Brussino, & Fernández-Dols, 2013).

There is some evidence showing that people may experience less intense emotional reactions to aversive stimuli in a FL (e.g., Dewaele, 2004; Harris, Ayçiçeği, & Gleason, 2003; Pavlenko, 2005; Puntoni, de Langhe, & van Osselaer, 2009). For example, childhood reprimands elicit decreased physiological responses in a FL versus a NL (e.g., Harris et al., 2003) and this may be dependent on the age of acquisition (e.g., Harris, 2004). This is because there are at least two major differences between the acquisition of a NL and a FL: the amount of exposure and the contexts of use. NL words benefit from more years of exposure and typically occur in more varied and emotional contexts. Even with equal proficiency and maximum use of two languages, the NL is perceived as emotionally stronger and is preferred for emotional expression (Dewaele, 2011). That is, though proficient FL users may understand the semantics of emotional words, they may not feel their full effect (for the emotional contexts of learning theory, see Harris, Gleason, & Ayçiçeği, 2006).

The difference in the contexts of acquisition between a NL and a FL may also lead to the latter being less sensorily or emotionally embodied (e.g., Pavlenko, 2012). Because FLs are typically not learned in naturalistic contexts, words may not occur in conjunction with affective states. FL classrooms do not provide sufficient opportunity for socialization or grounding the terms in bodily experience, and this may reduce the emotional impact of FL words (e.g., Dewaele, 2004; Pavlenko, 2004). Furthermore, because of the differences in acquisition between a NL and a FL, FL words are processed less automatically (Colbeck & Bowers, 2012; Segalowitz, Trofimovich, Gatbonton, & Sokolovskaya, 2008; Winkler, 2013; but see Eilola, Havelka, & Sharma, 2007; Sutton, Altarriba, Gianico, & Basnight-Brown, 2007). However, even when a FL is just as automatic in terms of processing aversive stimuli, the physiological response to these stimuli is larger in the NL (Eilola & Havelka, 2011).

It is also possible that a FL does not reduce overall emotionality, but selectively affects negative or positive affect (Wu & Thierry, 2012; Hadjichristidis, Geipel, & Savadori, 2015). Indeed, there is some evidence that negative words, but not positive ones, are less embodied in a FL than in a NL (e.g., Foroni, 2015; Sheikh & Titone, 2016). This suggests that in a FL, there may be less emotional conflict regarding negative content or a higher preference for the positive over the negative. Furthermore, the use of a FL dampens the emotional salience of the self (Ivaz, Costa, & Duñabeitia, 2016). This would predict that, when using a FL compared with a NL, people would be less inclined to base their choices on self-relevant factors, perhaps resulting in a more emotionally distant decision process.

This is consistent with a related explanation for the effect of language: Using a FL might create more psychological distance than using a NL (Costa et al., 2014a; Keysar, Hayakawa, & An, 2012). Indeed, increasing psychological distance seems to increase utilitarian preferences by shifting focus from the means, the sacrifice of one, to the goal or the outcome, saving five. For instance,

Aguilar et al. (2013) demonstrated that considering moral dilemmas from a more distant perspective leads to greater willingness to sacrifice for the greater good, even when this choice has negative consequences for oneself. Hence, if using a FL increases psychological distance, it may prompt potential gains, such as five lives, to be more salient or important than potential losses, such as one life. Indeed, the reduced emotional response and psychological distance accounts are not mutually exclusive. A more distant perspective could lead to a decrease in affective processing given the reduced salience of potential losses. This is congruent with the findings regarding the FLE on economic decisions; when using a FL, people are less likely to base their choices on avoidance of negative outcomes and are less susceptible to the effects of prospective loss, risk, and uncertainty than when using a NL (Costa, Foucart, Arnon, Aparici, & Apestequia, 2014; Keysar et al., 2012). Given all the ways in which using a FL may affect the manner in which people consider moral dilemmas, it is important to assess whether choices are affected by the same factors in FL contexts as in NL ones.

Overview of the Studies

We report nine experiments that evaluate the robustness of the FLE and various factors that affect it. This has allowed us to better describe the phenomenon and investigate potential explanations. The experiments are organized in three sets. In the first set of studies (Experiments 1a and 1b), we replicate the phenomenon directly and conceptually. We also test and rule out the alternative explanation that the effect of language might result from random responding. In the second set of studies, we assessed two additional alternative accounts. Experiment 2a tested whether increased cognitive control that results from language switching can account for the increase utilitarian choice. Experiment 2b evaluated whether using a FL prompts different social inferences about the group membership of the actors. In the third set of studies (Experiments 3a through 3e), we assessed the roles that action and consequences, as well as focus on them, play in the FLE. This was to assess if these factors affect choice differently in NL and FL contexts, which would be consistent with our proposed explanation. Finally, we also investigated whether proficiency level has a reliable effect on the phenomenon. Together, these investigations provide some explanation for the phenomenon as well as delineating its boundaries.

Experiments: General Method

Although the experiments differed in some aspects, they also shared many methodological details.

Participants, Materials, and Procedure

For each experiment, we collected samples of around 100 participants per language condition. We gathered the data from groups of 10 to 50 students attending university classes of varying majors, such as education, linguistics, literature, engineering, and psychology. The classes took place at universities in Spain (Universitat Pompeu Fabra, Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat de Girona, Universitat Rovira i Virgili, and Universitat Jaume I). Participants were late learners of English,

which was typically their second language after Spanish or third after Spanish and Catalan.³ To have relatively homogeneous samples, participants were excluded if they had lived in an English-speaking country for more than 12 months, had a NL that was not native to Spain, were under 18 years or over 40 years of age, did not complete the entire survey, had previously encountered the dilemmas, or reported understanding less than 50% of the dilemmas (in the FL condition). See Table 1 for the number of participants excluded from each study.

All materials were translated from English into Spanish, approved by multiple native speakers of Spanish, and back-translated for comparability (Brislin, 1970). For all experiments, each group of participants was randomly assigned to a language condition. Hence, participants received the instructions and materials in only one language (except for Experiment 2a). Each participant was presented with two dilemmas and the order of presentation was counterbalanced (except for Experiment 2a). The participants were reassured that there were no right or wrong answers, asked to respond in the order that the questions appeared, and to not change their answers once they had responded. Participants were asked to respond *yes* or *no* to each dilemma. All participants were asked for their age, NL(s), and gender. Within each experiment, the NL and FL groups had comparable ages and gender distributions.

In each experiment, we presented participants with both the footbridge and switch dilemmas. As we have seen in previous studies, the use of a FL seems to affect choice in the footbridge dilemma, but not in the switch dilemma. Hence, the switch dilemma in the current studies served as a control condition.

After responding to both dilemmas, participants assigned to the FL condition provided language background information. This included what percentage of the dilemmas they comprehended; the age of first exposure to English; number of months spent in an English-speaking country; and self-reported proficiency in reading, writing, speaking, and comprehension using a Likert-like scale ranging from 1 (*low proficiency*) to 7 (*high proficiency*). This information can be found in Table 2. The experiments lasted approximately 15 min and the experimenter remained in the classroom for the duration of the experiment.

Experiment 1: Replications and the Role of Vocabulary Knowledge

These studies aimed at replicating the original FLE and assessing the contribution of language knowledge.

Experiment 1a. Exact Replication and the Role of Vocabulary Knowledge

Experiment 1a was conducted as an exact replication of Costa et al.'s (2014a) study. Participants were presented with the switch and footbridge dilemmas in either their NL (Spanish) or a FL (English). We also investigated if FL vocabulary knowledge about key words in the dilemmas (ability to correctly translate them) modulates the effect. If the effect persists even when those using their FL possess sufficient vocabulary knowledge to understand

³ Knowledge of Catalan did not affect participants' responses in either the Spanish (NL) or English (FL) conditions.

Table 1
Number of Participants Excluded in Each of the Experiments for Either Comprehension/Demographic Problems or for Having Previous Exposure to the Experimental Dilemmas

Experiment	Demographic/comprehension exclusions	Previous experience with dilemmas	Total exclusions
Exp 1a	28	49	77
Exp 1b	3	0	3
Exp 2a	17	0	17
Exp 2b	19	38	57
Exp 3a	10	0	10
Exp 3b	12	0	12
Exp 3c	3	0	3
Exp 3d	18	107	125
Exp 3e	22	0	22

the dilemmas, this would suggest that it is not driven by a lack of comprehension leading to random responding.

Method

Participants, materials, and procedure. Two hundred eleven participants were included in Experiment 1a. Of the participants, 105 (64% female) participated in their NL and 106 (81% female) in a FL. The dilemmas presented were as follows:

The switch dilemma. A train is going down a track very fast toward five people. The train has a problem and cannot be stopped. Five people will die if you stay on this track. There is another track that you can use to divert the train. At the end of this track there is one man that will die if you change the track. Would you change the track?

The footbridge dilemma. A train is going down a track very fast toward five people. The train has a problem and cannot be stopped unless a heavy weight is dropped on the track. There is a very fat man next to you—your only way to stop the train is to push him onto the track, killing him to save five people. Would you push him?

After responding to the moral dilemmas, we gauged the relevant vocabulary knowledge in English of those in the foreign language condition by asking participants to translate 10 key words or phrases into Spanish (*toward, change the track, five people, divert, stay, heavy weight, drop, push, kill, save*). These terms were chosen as they pertain to a variety of aspects of the dilemmas such as the details of the utilitarian actions required in each dilemma, the consequences, and the general content. We then constructed a measure of language knowledge by each participant's translation score, calculating the total number of words/phrases they correctly translated out of the possible 10.

Results and Discussion

There was a significant effect of language on choice in the footbridge dilemma such that participants who used a FL made significantly more choices to save the larger number of people than those who used their NL (FL: 43% vs. NL: 19%), $\chi^2(1, N = 211) = 14.55, p < .001$. Unlike previous studies, this was also the case for the switch dilemma (FL: 87% and NL: 74%), $\chi^2(1, N = 211) = 5.27, p = .022$.

We split the responses from the FL group in two groups according to the median translation score in the vocabulary test: (a) low level = participants with translation scores lower than 8 (out of 10) and (b) high level = participants with translation scores higher than 8. None of the participants scored lower than 5.⁴

There was a significant difference between choices made by the low- and high-level groups for the footbridge dilemma such that the low-level group was more likely to choose to save the larger number of people (low-level group: 56% vs. high-level group: 33%), $\chi^2(1, N = 104) = 5.61, p = .018$. There was no significant difference between the low and high-level groups in choices for the switch dilemma (low-level group: 90% vs. high-level group: 83%), $\chi^2(1, N = 104) = 1.32, p = .25$. Importantly, both FL groups were made significantly more choices to save the larger number of people than the NL group (19%) in the footbridge dilemma (low-level group vs. native: $\chi^2[1, N = 157] = 21.84, p < .001$; high-level group vs. native: $\chi^2[1, N = 157] = 3.59, p = .058$). For switch, the difference was significant only for the low level group (NL: 74%), $\chi^2(1, N = 157) = 5.55, p = .018$.

Another way to assess the effect of FL proficiency is to split participants (median) according to their self-reported proficiency measured by (a) the percentage of understanding of the dilemmas reported and (b) the average score in the four abilities (reading, writing, speaking, and listening). Although both of these measures of self-reported proficiency were included in the analyses, the former value was given priority as one's subjective sense of having understood the particular text at hand seems more likely to gauge relevant proficiency than if they feel proficient in their FL in general. Here, no differences between the low level and the high level groups were observed (footbridge: 45% vs. 42%; switch: 89% vs. 85%).

Finally, we assessed whether the frequency of utilitarian responses to footbridge increases as FL proficiency decreases. We conducted a logistic regression treating translation score and av-

⁴ In order to determine the effect of relevant language knowledge, we further pruned the data to include only those participants who correctly translated at least three of the 5 terms that were essential for understanding the dilemmas ("towards", "change the track", "push", "kill" and "save"; this led to the exclusion of 2 participants). This was to ensure sufficient comprehension of the dilemmas.

Table 2

Demographic and Background Information for all Participants in Foreign Language Conditions Including, Age of First English Exposure, Months Spent Abroad in an English-Speaking Country, Self-Rated English Proficiency (1 = low proficiency, 7 = high proficiency) for Reading, Writing, Listening, and Speaking, and Average Percentage of Foreign Experiment Materials Understood

Experiment	Age of exposure	Months abroad	Reading	Writing	Listening	Speaking	Percentage understood
Exp 1a	6.5	<1	5.4	4.9	5.4	4.5	87
Exp 1b	6.1	<1	5.2	4.4	5.3	4.1	92
Exp 2a	6.8	<1	5.2	4.5	5.1	4.1	88
Exp 2b	7.1	<1	5.5	4.8	5.4	4.6	92
Exp 3a	6.9	<1	4.9	4.4	4.9	3.9	80
Exp 3b	6.8	<1	5.1	4.4	5.2	4.0	85
Exp 3c	6.5	<1	5.6	4.8	5.5	4.4	91
Exp 3d	6.5	1	5.6	5.0	5.5	4.6	93
Exp 3e	6.8	<1	5.1	4.5	5.2	4.0	89

erage self-reported proficiency in the four abilities as continuous variables to evaluate if they significantly predict choice. A test of the full model against a constant only model was not significant, indicating that including proficiency measures did not improve the model, $\chi^2(2) = 3.01, p = .22$. Although the Exp(B) values suggest that every decreased unit of proficiency increases the odds ratio of making the utilitarian choice for both translation score (Exp[B] = .89) and average self-reported proficiency (Exp[B] = .78), neither predictor was significant (translation: $\beta = -0.11, SE = .14, p = .43$; reported proficiency: $\beta = -0.25, SE = .23, p = .28$).

In this experiment, we replicated the FLE for the footbridge dilemma: People were more willing to choose the utilitarian choice when using their FL compared with when using their NL. We also found that the phenomenon was present when people had high vocabulary knowledge. This suggests that the origin of the phenomenon cannot be entirely attributed to a lack of understanding.

Experiment 1b: Conceptual Replication With Different Dilemmas

Thus far, most published studies that demonstrated the FLE on responses to moral dilemmas have only used the footbridge and switch versions of the trolley problem (although see Geipel et al., 2015; Study 3). To assess the generalizability of the pattern—effect of language for footbridge-like but not switch-like dilemmas—in this experiment, we conducted a conceptual replication using the following dilemmas:

The Hospital Dilemma (Adapted From Thomson, 1985)

You are working in a hospital. There is a fire in the street, and there is smoke coming through the hospital's ventilation system. In one room of the hospital there are five patients. In another room there is only one patient. If you do nothing the smoke will go into the room with the five patients and they will die. There is a button that diverts the ventilation system. If you push the button, the smoke will go into the room with one patient and the patient will die, but the five patients in the other room will be safe. Would you push the button?

The Terrorist Dilemma (Adapted From Greene et al., 2001)

You are negotiating with terrorists to save a group of six tourists that have been captured. The leader of the terrorists gives you the

choice: If you choose one tourist and shoot him, the other five tourists will be safe; if you decide not to kill anybody, the terrorist will kill five tourists and one will be safe. Would you kill one tourist?

These dilemmas are conceptually similar to switch and footbridge, respectively. The hospital dilemma is similar to the switch dilemma in that sacrificing the one person does not require direct personal force and the death of that person could be interpreted as a side effect of saving five people. Given that the effect of language tends to be minimal when dealing with these less personal scenarios, we did not expect to find the FLE for the hospital dilemma. On the other hand, the terrorist dilemma is similar to footbridge as it requires direct personal force to kill one person, a death that is instrumental to saving the other five; hence, for this dilemma we expected to find an effect of language.

In our view, a reduction in emotional response and an increase in psychological distance should lead those using a FL to be less sensitive to variations such as the use of personal force and the instrumentality of the one person's death. A reduction in emotional response in particular for negative and self-relevant emotions should reduce the aversion to using personal force and the person's death being instrumental. In a similar vein, an increase in psychological distance would limit the weight of personal involvement (force) and the means (an instrumental murder) relative to the consequences. Therefore, we expect to replicate the pattern previously found for conceptually similar dilemmas, which would predict an effect of language for terrorist but not hospital.

Method

Participants, materials, and procedure. In Experiment 1b, 173 participants were included. Ninety-three (84% female) participated in their NL (Spanish) and 80 (84% female) in a FL (English). The same materials and procedure were used as in Experiment 1a.

Results and Discussion

There was a significant effect of language on choice in the terrorist dilemma such that participants who used a FL made significantly more choices to save the larger number of people than did those who used their NL (FL: 54% vs. NL: 39%), $\chi^2(1, N = 173) = 3.92, p = .048$. This difference was not significant for the

hospital dilemma (FL: 83% vs. NL: 73), $\chi^2(1, N = 173) = 2.17$, $p = .14$.

Following the same procedure as in Experiment 1a, we also investigated the effect of self-reported FL proficiency. Here, proficiency did not significantly affect people's choices (terrorist low level: 55% vs. high level: 52.5%; hospital low level: 80% vs. high level: 85%). In this experiment, we replicated the pattern previously found: There was a FLe for terrorist but not hospital, as expected given the effect for footbridge and not switch.

Experiment 2: Language Switching and Social Inferences

In these two experiments, we investigated the roles of language switching and social inferences in the presence of the FLe.

Experiment 2a: Is the FLe a Consequence of Language Switching?

It has been argued that the FLe could be driven by language switching rather than by FL use per se (Oganian, Korn, & Heekeren, 2016). The argument goes as follows: Because participants presumably use the NL in normal circumstances, when they are confronted with dilemmas in their FL, they are required to switch languages. This switching activity may lead to an increase in cognitive control that may in turn affect the way the dilemma is processed. Indeed, independent evidence suggests that increasing cognitive control could be associated with an increase in utilitarian choices (see the introductory paragraphs). Hence, it is possible that the FLe is no more than a switching effect in disguise.

If this is correct, then asking all participants to switch languages before responding to the footbridge dilemma should attenuate or remove the effect of language. The reasoning is the following: If participants are presented with the footbridge dilemma in the NL preceded by the switch dilemma in the FL, utilitarian choices in footbridge should increase for this group as a result of language switching, thus limiting the impact of language. To assess this issue, we first presented participants with the switch dilemma, in either their NL or FL and then evaluated responses to the footbridge dilemma in the other language. If language switching is what causes the effect in the footbridge dilemma, then it should not be present here. However, if the effect is not a consequence of language switching, but depends on the language of presentation of the key dilemma, then the phenomenon should persist.

Method

Participants, materials, and procedure. Two hundred four participants were included in Experiment 3. One hundred one (85% female) participated using their NL (Spanish) for footbridge (NL footbridge), and 103 (75% female) participated using their FL (English) for footbridge (FL footbridge).

Unlike previous experiments, here all participants received the instructions in their NL. Participants then received the switch dilemma followed by the footbridge dilemma. The dilemmas are the same as those in Experiment 1a. Crucially, we manipulated the languages to examine the effect of language switching. Those in the FL footbridge condition received the switch dilemma in their NL and then the footbridge dilemma in their FL. Those in the NL

footbridge condition received the switch dilemma in their FL and then the footbridge in their NL.

Results and Discussion

There was a significant effect of language on choice in the footbridge dilemma such that participants who used their FL made significantly more choices to save the larger number of people than those who used their NL for the footbridge dilemma (FL: 47% vs. NL: 29%), $\chi^2(1, N = 204) = 6.95$, $p = .008$. There was no effect of language on responses to switch (FL: 84% vs. NL: 91%), $\chi^2(1, N = 204) = 2.39$, $p = .12$.

We also investigated the effect of self-reported FL proficiency. We only evaluated choices based on information presented in their FL (responses to footbridge for Foreign footbridge and responses to switch for NL footbridge). For footbridge, the effect of proficiency was not significant (low level: 54% vs. high level: 39%), $\chi^2 = (1, N = 103) = 2.22$, $p = .14$. For switch, there was an effect of proficiency (low level: 75% vs. high level: 94%), $\chi^2(1, N = 101) = 7.19$, $p = .007$.⁵

In this experiment, even though all participants switched languages, we found that using a FL for footbridge still lead to more choices to save the larger number of people (18%) than using the NL. Indeed, the magnitude of the FLe was similar when no language switch was involved (Experiment 1a; 24%).

To directly evaluate the possibility that language switching may diminish the effect of language, we compared the results of the footbridge dilemma from Experiment 2a with those from Experiment 1a. We did this by fitting a logistic regression with two between-subjects bivalent factors (Experiment: 1a vs. 2a \times Language: native vs. foreign). If having all participants switch languages reduces the impact of language, then we should find a significant interaction between these two factors.

The results showed a main effect of language (Wald statistic = 7.76, $p = .005$), but not of language context (experiment: Wald statistic = 2.62, $p = .105$) or of the interaction (Wald statistic = 0.887, $p = .346$). That is, although numerically speaking the magnitude of the effect was larger when not all participants were required to switch languages (24%) than when they were (18%), these variations were not significantly different from one another.

In this experiment, we showed that the FLe was present when all participants are required to switch languages. Moreover, the comparison of the results of this experiment to those of Experiment 1a revealed that requiring all participants to switch languages does not significantly reduce the FLe. Hence, the presence of the phenomenon cannot be attributed to an increase in cognitive control associated with language switching.

Experiment 2b: Is the FLe a Consequence of Social Inferences?

In this experiment we assessed the potential role that social inferences may have in the presence of the FLe. It is possible that the use of the NL may prompt participants to infer that the actors

⁵ When compared with the NL group (91% utilitarian choices) for the switch dilemma, the low level group differed significantly, $\chi^2(1, N = 154) = 7.82$, $p = .005$, but the high level group did not, $\chi^2(1, N = 153) = 0.35$, $p = .55$.

in the dilemmas are in-group members whereas the use of their FL may prompt participants to infer that they are out-group members. These different attributions may contribute to the effect by altering how participants relate to the actors and, consequently, the value these lives are given. Indeed, independent evidence reveals that the in-group or out-group status of the actors involved in a moral dilemma can affect the decisions that people make (e.g., Swann, Gomez, Dovidio, Hart, & Jetten, 2010).

To evaluate this issue, the dilemmas included in Experiment 2b explicitly stated the group membership of the five victims such that they were the same for both language groups. In addition, this experiment varied group membership, specifying that the five victims were either in-group (Spanish) or out-group (American). If the impact of a FL is contingent on different group membership inferences by those using one language or another, then making group membership explicitly the same would undo the effect. On the other hand, if the effect persists, this suggests that it cannot be attributed to different social inferences.

Method

Participants, materials, and procedure. Three hundred ninety-nine participants were included in Experiment 2b. They were randomly assigned to either the in-group or out-group condition and to participate in either their NL (Spanish) or FL (English). Hence, there were four groups of participants in this experiment. In the in-group conditions, 95 (69% female) participated in their NL and 99 (71% female) in their FL. In the out-group conditions, 103 (81% female) participated in their NL and 102 (84% female) in their FL.

Participants in the in-group conditions were told that the five people were Spanish whereas those in the out-group conditions were told that the five people were American. For both the switch and footbridge dilemmas, participants received the exact scenarios used in Experiments 1a and 2a, except that the first line was adjusted to identify the five victims' nationalities as in-group or out-group in the following manner: "A train is going down a track very fast toward five Spanish [American] people."

Results and Discussion

There was a significant effect of language on choice in the footbridge dilemma such that participants who used their FL made significantly more choices to save the larger number of people than those who used their NL. This was the case both for the in-group condition (FL: 27% vs. NL: 15%), $\chi^2(1, N = 194) = 4.57, p = .033$, and the out-group condition (FL: 36% vs. NL: 19%), $\chi^2(1, N = 205) = 7.26, p = .007$. There was no effect of language on responses to switch (in-group: FL: 72% vs. NL: 75%; $\chi^2[1, N = 194] = 0.23, p = .64$; out-group: FL: 80% vs. NL: 78%; $\chi^2[1, N = 205] = 0.23, p = .63$).

We also investigated the effect of self-reported FL proficiency. There were no significant differences between the low-level and high-level groups (footbridge in-group: 30% vs. 24%; footbridge out-group: 33% vs. 39%; switch in-group: 74% vs. 69%; switch out-group: 76% vs. 84%), and all FL groups gave more utilitarian responses than did the NL groups in footbridge.

In this experiment, we showed that the effect of language was present when the group membership of the actors in the dilemma

was specified. However, it may be that the effect of language was larger when the victims were out-group members (17%) compared with when they were in-group ones (12%).

To more directly evaluate whether the effect of group membership (in- vs. out-group) interacts with the effect of language, we fitted a logistic regression with two between-subjects bivalent factors (Social Context: in-group vs. out-group \times Language: native vs. foreign). If the victims' social group affects the FLe, then we should find an interaction between these two factors. However, the results showed that they do not interact (Wald statistic = 0.030, $p = .862$). That is, the size of the effect of language was the same regardless of whether the victims were in-group or out-group members.

In this experiment, we observed the FLe despite specifying the in-group/out-group status of the actors involved in the dilemma. Hence, this suggests that the FLe cannot be explained by different social inferences, given that the effect remains present when the dilemmas does not allow for such inferences. Furthermore, the effect of language was the same whether the victims were in-group or out-group.

Thus far the current studies have provided evidence that the moral FLe is robust and present in various contexts. They have also shown that the effect cannot be attributed to (a) a lack of understanding associated with FL processing, (b) an increase in cognitive control associated with language switching, or (c) different social inferences prompted by the language of presentation. In the next five experiments, we addressed further factors that may affect the presence of the phenomenon, paying special attention to how actions and consequences modulate its magnitude.

Experiment 3: Actions, Consequences, and Focus

The first three studies of this section addressed the role of the type of action involved in the dilemma and the way that the choice is framed, in the presence of the FLe. The reasoning for this was the following: It is possible that the FLe is a result of reducing the aversion elicited by the action required to save the larger number of lives. In the footbridge dilemma, pushing the man off the bridge is by itself an action that elicits aversion regardless of its outcome. Indeed, independent research has shown that people tend to experience high levels of aversion to performing typically violent actions even in mock situations that do not lead to any harmful consequences (pulling the trigger of a toy gun; Cushman et al., 2012; see also Miller & Cushman, 2013; Miller, Hannikainen, & Cushman, 2014). As these authors argue, this aversion may be one of the main reasons that people avoid choosing to sacrifice a person when it involves violent action. Furthermore, some claim that intuitions differ for switch and footbridge because of the type of action involved and whether or not we have experience with it leading to negative outcomes (for a discussion of model-free vs. model-based control, see Cushman, 2013; Railton, 2016). Therefore, it is reasonable to think that any factor that reduces action aversion could lead to an increase in utilitarian choices.

Using a FL could be one of them; as we have argued, there is evidence suggesting that the use of a FL decreases emotional response and increases psychological distance relative to that of a NL. Importantly, both of these factors would predict a decrease in action aversion and an increase in choosing to sacrifice the one to save the larger number of people. First, a reduction in emotional

response even in healthy individuals leads to this, presumably by reducing social cognition e.g., empathy (Duke & Bègue, 2015). Second, it has been shown that inducing psychological distance leads to an increase in willingness to sacrifice for the greater good (Aguilar et al., 2013), presumably by increasing the focus on outcomes (ends) as opposed to actions (means). Hence, the use of a FL could lead to a reduction in the importance of the means or an increase in the importance of the consequences, altering the way the problem is considered. Indeed, there is recent evidence suggesting this is the case: Geipel and colleagues (2016) found that FL use reduces the weight of the means (intentions) and increases the weight of the outcomes in moral evaluations. Furthermore, the authors argue that this is consistent with a reduction in intuitive processes, which is congruent with there being an effect of language for footbridge but not switch.

To investigate these issues, in Experiments 3a and 3b we assessed whether the FLE is a result of reduced action aversion by (a) making the action less aversive by removing personal force and (b) framing the choice in terms of the consequences of inaction, which should limit the potential contribution of action aversion. Experiment 3c investigated the salience of the trade-off between the means and the consequences in the FLE. Specifically, whether the FLE would persist when this trade-off is made highly salient. Given that the use of a FL may prompt more focus on the consequences than on the means required to achieve it, making the trade-off very explicit may reduce the impact of language.

The remaining two studies investigated the role of the negative consequences associated with saving the larger group. Reduced emotional response and increased psychological distance would lead to a reduction in (a) the salience or importance given to self-relevant emotions and (b) the sensitivity to the negative relative to the positive consequences of action. Therefore, Experiments 3d and 3e investigated if the severity of the negative consequences has less weight in choices made in FL contexts than in NL ones. We explain this reasoning in more detail in the introduction of these experiments.

Experiment 3a: Reducing Action Aversion

In this experiment, we assessed whether the FLE would be present when the action needed to save the larger number of people is less aversive (pushing a button instead of pushing a man). The idea was that if the FLE is driven by reduced action aversion, then it should be attenuated or absent when the action involved in the dilemma elicits less aversion.

To evaluate this hypothesis, we used a version of the footbridge dilemma (adapted from Greene et al., 2009) in which saving the larger number of people does not involve physically pushing the man to his death. Instead, it requires a rather neutral action: pushing a button. However, the consequence of this action is that the man falls onto the track, stops the trolley and dies, and this leads to saving five people. Although the consequence of taking action was the same as in the standard footbridge dilemma, this “button” version did not require physical contact, personal force, or a typically negative action. Indeed, pushing a button is not intrinsically associated with negative consequences, but pushing a man is. The button dilemma still differed from switch in that a man’s death is instrumental, whereas in the switch dilemma this death can be construed as a side effect. Therefore, this experiment

also allowed us to assess the role of personal force in the phenomenon. If the effect of the FL is the result of a reduced aversion to this factor, then the phenomenon should be reduced.

Method

Participants, material, and procedure. Two hundred two participants were included in Experiment 3a. One hundred (68% female) participated in their NL (Spanish) and 102 (78% female) in their FL (English). Participants received the switch (as in previous experiments) and button dilemmas. For the button dilemma, participants received the following scenario:

A train is going down a track very fast toward five people. The train has a problem and cannot be stopped, unless a heavy weight is dropped on the track. There is a very fat man next to the track—your only way to stop the train is to push a button that will make him fall onto the track, killing him to save five people. Would you push the button?

Results and Discussion

Using the FL led to more choices to save the larger number of people in button than did using the NL, and this difference was marginally significant (FL: 67% vs. NL: 55%), $\chi^2(1, N = 202) = 2.89, p = .089$. There was no effect of language on responses to the switch dilemma (FL: 75% vs. NL: 78%), $\chi^2(1, N = 202) = 0.18, p = .67$.

We also investigated the effect of self-reported FL proficiency. There was no effect of proficiency for the button dilemma (low level: 67% vs. high level: 67%). There was an effect of proficiency for switch (low level: 63% vs. high level: 88%), $\chi^2 = (1, N = 102) = 8.96, p = .003$.⁶

This experiment showed a clear trend toward the presence of a FLE in the button dilemma. However, it failed to reach conventional significance. A further way to assess whether the effect is driven by reduced action aversion is to compare its magnitude when the action required to achieve the utilitarian goal involved personal force and a typically violent action (pushing the man; Experiment 1a) to when it did not (pushing the button; Experiment 3a). We did this by fitting a logistic regression with two between-subjects bivalent factors (Experiment: 1a vs. 3a \times Language: native vs. foreign).

The results showed significant main effects of both factors (experiment: Wald statistic = 26.57, $p < .001$; language: Wald statistic = 11.78, $p = .001$), but the interaction was not significant (Wald statistic = 2.57, $p = .11$). Thus, it appears that the effect of language was similar in these two contexts.

To further explore the potential difference in the effect sizes found for the footbridge dilemma between Experiments 1a and 3a, we calculated odds ratios as a means to quantify the associations between language and utilitarian choice. Here they represent the odds that choosing the utilitarian option will occur given the use of the FL, compared with the odds of choosing it given the use of the NL (see Szumilas, 2010). If the odds ratio is >1 , then choosing the

⁶ For the switch dilemma, compared with those who used the NL (78% utilitarian choices), the low level group differed significantly (63%), $\chi^2(1, N = 151) = 3.97, p = .046$, but the high-level group did not (88%), $\chi^2(1, N = 151) = 2.34, p = .13$.

utilitarian option is considered to be more likely given the use of the FL.

At a descriptive level, the odds ratios between language and choice were ordered as expected: 3.258 and 1.636 for Experiments 1a and 3a, respectively. This suggests that the FLe may have been larger when the action required to save the larger number of people was highly aversive (pushing a man) than when it was not (pushing a button). In addition, the increase in choosing to act in Experiment 3a compared with Experiment 1a was numerically larger for NL users (36%) than for FL users (24%). This may suggest that reducing action aversion had a larger effect on those using the NL. Thus, the effect of reducing action aversion may have had a larger impact on those using their NL relative to those using their FL; however, the lack of a significant interaction in the logistic regression precludes us from definitively drawing this conclusion.⁷

Thus, we found that the magnitude of the FLe was statistically similar whether the action that leads to saving the larger number of people involved personal force or not. Hence, it appears that the phenomenon cannot be attributed, solely, to the potential reduction in action aversion that the use of a FL may confer. However, our conclusions are tentative given that both the main effect for Experiment 3a and the interaction between language and experiment (1a vs. 3a) are arguably marginal. Nevertheless, in numerical terms, the magnitude of the phenomenon was larger when the action involved personal force, which leaves open the possibility that action aversion may contribute to it to some degree. In the next experiment, we further explored this issue from a different perspective.

Experiment 3b. The Effect of Asking About Actions or Consequences

As described in the preceding text, the FLe may come about because of a difference in the way the dilemma is considered by each language group. We have argued that a FL context might reduce action aversion or lessen its importance relative to the consequences (as compared with a NL context). If this is the case, it would reduce the impact of action aversion and subsequently favor choosing to sacrifice the one in footbridge. Indeed, in the previous experiment, we observed a reduction (numerically speaking) in the magnitude of the FLe when action aversion was presumably reduced. Another way to assess the potential contribution of action aversion is to frame the dilemmas' choices such that they do not focus on taking action. This can be achieved by stressing the consequences of inaction (five people will die) rather than by focusing on the action itself (e.g., pushing the man), when presenting the choices. Previous research shows that similar manipulations lead to an increase in utilitarian choices (Petrinovich & O'Neill, 1996). After all, intuitively, it seems more acceptable to say "No" to "Would you let five people die?" (or "Yes" to "Would you save the five people?"), rather than "Yes" to "Would you push the man?"

In this experiment, we asked the following question after the footbridge dilemma: "Would you let five people die?" instead of "Would you push the man?" Creating conditions that, presumably, would reduce the impact of action aversion on choice allows us to further explore its potential role in the FLe. Specifically, this increases the salience of the cost associated with inaction and makes the choice to save the larger number more intuitive. This

manipulation should increase willingness save the larger number in the footbridge dilemma (but not in switch; see Broeders, Van Den Bos, Müller, & Ham, 2011). If reduced action aversion is the main cause of the FLe, then it should not be present here as action aversion would be minimal.

Method

Participants, material, and procedure. One hundred ninety participants were included in Experiment 3b. Ninety-seven (69% female) participated in their NL (Spanish) and 93 (72% female) in their FL (English).

Participants received the same switch ("consequence switch") and footbridge ("consequence footbridge") dilemmas as in Experiments 1a. Unlike before, rather than asking participants whether they would "change the track" or "push him," they were asked whether they would "let five people die." Also unlike before, here the same question was posed for both dilemmas.

Results and Discussion

There was a significant effect of language on choice in consequence footbridge such that those who used their FL made significantly more choices to not let five people die than those who used their NL (FL: 60% vs. NL: 41%), $\chi^2(1, N = 190) = 6.84, p = .009$. There was no effect of language for the consequence switch dilemma (FL: 83% vs. NL: 84%), $\chi^2(1, N = 190) = 0.017, p = .90$.

We also investigated the effect of self-reported FL proficiency. There were no significant effects (consequence footbridge low level: 53% vs. high level: 67%; $\chi^2[1, N = 93] = 1.96, p = .16$; consequence switch low level: 83% vs. high level: 83%).

The results of this experiment revealed that the FLe was also present when the dilemmas are framed in terms of the consequences of inaction, which would presumably reduce the potential effects of action aversion. This suggests that the phenomenon is indeed robust as it is not contingent on questions about action or affirmative responses. Hence, although a reduction in action aversion may contribute to the presence of the FLe, the current results suggest that it cannot fully account for it.

Experiment 3c. The Trade-Off Between the Means and the Consequences

In this experiment, we further assessed potential differences in the way the dilemma is considered depending on the language. We did so by focusing attention on the trade-off between the means and the consequences posed by the dilemma. As we have argued, people may give different importance to actions and consequences

⁷ When the previous analyses were repeated for the switch dilemma, the results were different. There was a significant main effect of language (Wald statistic = 5.28, $p = .022$) but not of experiment (Wald statistic = 0.39, $p = .53$). There was also a significant interaction (Wald statistic = 3.81, $p = .051$). The main effect of language was due to the fact that globally speaking there were more utilitarian choices made by those using the foreign than the NL. This appears to be primarily driven by Experiment 1a in which FL users were significantly more utilitarian, resulting in the interaction.

in their responses depending on the language context. Indeed, it is possible that the use of a FL prompts more focus on the consequences than on the trade-offs that one has to make to achieve them. In this context, making the trade-offs very explicit by focusing attention both on the consequences and on the means required may reduce the impact of language. We explored this issue by asking participants the following question after the footbridge dilemma: “Would you let five people die by not pushing him?”

Method

Participants, materials, and procedure. Two hundred one participants were included in Experiment 3c. One hundred three (46% female) participated in their NL (Spanish) and 98 (57% female) in the FL (English).

Participants received the footbridge and switch dilemmas as in Experiment 1a, but with different final questions. The “consequence action switch” dilemma ended with the question “Would you let five people die by not changing the track?” and the “consequence action footbridge” dilemma ended with the question “Would you let five people die by not pushing him?”

Results and Discussion

There was no significant difference in choosing to not let five people die for the consequence action footbridge dilemma between those who used their FL and NL (FL: 41% vs. NL: 38%), $\chi^2(1, N = 201) = 0.18, p = .67$. Similarly, there was no difference for the consequence action switch dilemma (FL: 76% vs. NL: 79%), $\chi^2(1, N = 201) = 0.28, p = .60$.

We also investigated the effect of self-reported FL proficiency. For consequence action footbridge, there was a significant effect of proficiency (low level: 53% vs. high level: 29%), $\chi^2(1, N = 98) = 6.08, p = .014$, but for consequence action switch there was not (low level: 82% vs. high level: 69%), $\chi^2(1, N = 98) = 1.99, p = .16$. For the consequence action footbridge dilemma, the low-level group chose marginally more often not to let five people die (53%) compared with the NL group (38%), $\chi^2(1, N = 152) = 3.13, p = .077$, but there was no effect of language for the high level group (29%), $\chi^2(1, N = 152) = 1.26, p = .26$. This experiment showed that the FLe disappeared when the choice posed by the dilemma stressed the trade-off between the consequences and the action required to achieve them.

In the last two experiments we have shown that the FLe was modulated by the wording of the question posed in the dilemma. The effect was strongest when action was made explicit (Experiment 1a: Would you push the man?), weaker when the consequences of inaction were made explicit (and the choice to save the larger number was more intuitive; Experiment 3b: Would you let five people die?), and null when the link between the two was made explicit (Experiment 3c: Would you let five people die by not pushing the man?).⁸ This pattern of results identifies the contribution of action aversion and an explicit focus on the trade-off have on the presence of the effect. In the following two experiments, we evaluated the way in which consequences contribute to the FLe.

Experiments 3d and 3e: The Negative Consequences Associated With Saving the Larger Group

In the following two experiments, we assessed the role that the consequences associated with saving the larger group have in the presence of the FLe. Part of the reluctance to opt for this choice in the footbridge dilemma could be due to the emotional reaction elicited by the negative consequences (being responsible for sacrificing a man’s life) that result from said choice. In other words, the idea of intentionally sacrificing the life of a person may elicit such a negative reaction (“I do not want to kill anyone”) that it prevents participants from accessing the logic behind the trade-off (one life for five lives). Reducing such aversion would allow for more cost-benefit analyses. Hence, if using a FL reduces the aversion elicited by the negative consequences of sacrificing the one, this could partially explain the effect.

If this is the case, then reducing the emotional reaction elicited by the consequences of acting may increase the inclination to choose this option and reduce the potential role that language may play. Arguably, a reduction in the degree of harm associated with sacrificing the one would lead to a reduction in the aversion elicited by said choice. This would, in turn, lead to an increase in choices to save the larger number of people. In fact, previous research has shown that decreasing the extent of harm increases willingness to choose this option in moral dilemmas when using a NL (Trémolière & De Neys, 2013). Imagine an extreme case: By pushing the man you save five lives, but he only suffers minor injuries such as a broken toe. Note that the action would be the same, pushing a man, but the consequences would likely elicit a much less intense emotional reaction. In such examples—or even in less extreme ones as those used in the following experiments—it is likely that the FLe would disappear, if indeed the phenomenon is basically attributable to a reduction of the emotional reaction to the negative consequences.

Said reaction could in part be attributable to self-relevant emotions about the negative consequences that making the utilitarian choice would have for one personally (e.g., the extent to which it

⁸To evaluate the effect of how the choice posed by the dilemma is framed, we compared the responses to the footbridge dilemma in three experiments as a function of the question (the only difference between the experiments): “Would you push the man?” (Experiment 1a), “Would you let five people die?” (Experiment 3b), and “Would you let five people die by not pushing the man?” (Experiment 3c). We did this by fitting a logistic regression with two between-subject factors (Experiment [1a, 3b, 3c] \times Language [native vs. foreign]). We expected to find the strongest association between language and decision when only action was made explicit, a moderate effect of language when only consequences were made explicit, but none when the trade-off between the two was made explicit. The results showed a significant main effect of question (Wald statistic = 8.316, $p = .004$) but not of language (Wald statistic = 0.49, $p = .49$). More importantly, we found a significant effect of the interaction between language and question (Wald statistic = 2.89, $p = .020$). The odds ratio between language and choice equaled 3.258, 2.157, and 1.132 for Experiments 1a, 3b, and 3c, respectively. This pattern was consistent with our expectations. When these analyses were repeated for the switch dilemma, the results were different. There was no main effect of any of the factors (experiment: Wald statistic = 0.60, $p = .44$; language: Wald statistic = 0.89, $p = .35$), but the effect of the interaction was significant (Wald statistic = 4.18, $p = .041$). This was probably due to the fact that we found a significant association between language and choices in Experiment 1a for the switch dilemma. This effect did not appear again in any of the other experiments and may be a Type I error.

is punishable). Failure to anticipate these self-relevant emotions has been linked with an increase in utilitarian choices (e.g., in individuals with brain lesions: Ciaramelli, Muccioli, Làdavas, & di Pellegrino, 2007; Moretto, Làdavas, Mattioli, & di Pellegrino, 2010). Hence, it may be that the use of a FL promotes willingness to make this choice by reducing the importance of these emotions (see Ivaz et al., 2016 for evidence suggesting that the self is less emotionally salient in a FL context compared with a NL one). If reducing the extent of harm associated with acting to save the larger number of people reduces the intensity of self-relevant emotions, we expect that this would have a larger effect on those using the NL. By increasing utilitarian choices for those using the NL, this would likely limit the effect of language.

Indeed, a reduction in sensitivity to negative consequences would also be consistent with another related explanation for the FLe. Perhaps, as we have argued, a FL context elicits psychological distance, compared with that of a NL. If this is the case, it may encourage participants to assess the trade-off in more general terms (harming one to save five) and therefore to dismiss, to a certain degree, how much harm is inflicted on the sacrificed person. As argued, an increase in psychological distance may also promote more focus on the positive than on the negative. This would be consistent with results showing that using a FL decreases negative affect and/or increases positive affect (Hadjichristidis et al., 2015; Wu & Thierry, 2012) and that negative, but not positive words, are less embodied in the FL context (Feroni, 2015; Sheikh & Titone, 2016). If this is the case, then variations in the negative consequences should have a smaller effect on those using a FL than on those using a NL.

Hence, if a FL reduces sensitivity to the negative consequences (or increases sensitivity to the positive ones) associated with the action required to save the larger number of people, the specific degree of harm associated with this action would be less impactful.

We assessed these issues by altering the footbridge dilemma such that the utilitarian choice would leave the fat man permanently disabled (Experiment 3d) or seriously injured (Experiment 3e).

Experiment 3d: Disability

Method

Participants, materials, and procedure. One hundred ninety-seven native Spanish speakers were included in Experiment 3d. Ninety-six (67% female) participated in their NL (Spanish) and 101 (67% female) in their FL (English).

Participants received the same switch dilemma as in previous experiments and a modified “disabled footbridge” dilemma. For the disabled footbridge dilemma, participants received the following scenario:

A train is going down a track very fast toward five people. The train has a problem and cannot be stopped, unless a heavy weight is dropped on the track. There is a very fat man next to you—your only way to stop the train is to push him onto the track. If you push the fat man, he will be disabled, unable to walk, for the rest of his life. Would you push him?

Results and Discussion

The FLe persisted for the disabled footbridge dilemma, $\chi^2(1, N = 197) = 5.54, p = .019$, with those using the FL more frequently choosing to save the larger number of people than those using the NL (FL: 40% vs. NL: 24%), $\chi^2(1, N = 197) = 5.54, p = .019$. There was no significant difference between languages for the switch dilemma (FL: 75% vs. NL: 79%), $\chi^2(1, N = 197) = 0.43, p = .51$.

We also investigated the effect of self-reported FL proficiency. There were no significant effects (disabled footbridge low level: 36% vs. high level 43%; switch low level: 73% vs. high level: 78%).

In this experiment, we observed that the FLe was present when the negative consequences of saving the larger number of people were somewhat less severe than death. This suggests that the phenomenon is not limited to cases of lethal harm.

Experiment 3e: Injury

Method

Participants, materials, and procedure. Two hundred twenty-three native Spanish speakers were included in Experiment 3e. Of these participants, 117 (83% female) participate in their NL (Spanish), and 106 (84% female) participated in their FL (English). Participants received the same switch dilemma as before and the modified “injured footbridge” dilemma. For the injured footbridge dilemma, participants received the following scenario:

A train is going down a track very fast toward five people. The train has a problem and cannot be stopped, unless a heavy weight is dropped on the track. There is a very fat man next to you—your only way to stop the train is to push the man onto the track, causing him serious injuries, to save these five people. Would you push him?

Results and Discussion

Those who used their FL made more choices to save the larger number of people than those who used their NL users for the injured footbridge dilemma but this difference was not statistically significant (FL: 43% vs. NL: 36%), $\chi^2(1, N = 223) = 1.31, p = .25$. There was no effect of language on choices made in switch (FL: 67% vs. NL: 64%), $\chi^2(1, N = 223) = 0.20, p = .65$.

We also investigated self-reported FL proficiency. There were no significant differences between groups (injured footbridge low level: 43% vs. high level: 43%; switch low level: 72% vs. high level: 62%).

In this experiment, we found that the FLe was absent when the negative consequences of saving the larger number of people were rather less severe and permanent than death. This reveals a potential boundary of the phenomenon, namely that the effect decreases when the negative consequences of taking action are diminished.

To more directly assess this issue, we compared the results for different footbridge dilemmas that vary in the degree of the harmful consequences associated with pushing the man: death (Experiment 1a), disability (Experiment 3d), and injury (Experiment 3e). We did this by fitting a logistic regression with two between-subjects factors (Experiments: 1a, 3d, 3e \times Language: native vs. foreign). We expected to find the strongest association between

language and choice when the severity of harm is highest (Experiment 1a), a weaker association when it is reduced (Experiment 3d), and the weakest association when it is the least severe (Experiment 3e).

The results showed a significant main effect of experiment (Wald statistic = 7.97, $p = .005$) but not of language (Wald statistic = 1.34, $p = .25$). More importantly, their interaction was significant (Wald statistic = 4.48, $p = .034$). The direction of the effect was as expected: The odds ratios between language and choices equaled 3.258, 2.081, and 1.369 for Experiments 1a, 3d, and 3e, respectively.⁹ Thus, this suggests that the effect of language is diminished when the negative consequences of the action associated with the utilitarian choice are made less severe.

Indeed, the frequencies of utilitarian choices changed significantly across these three experiments for the NL group, $\chi^2(2, N = 318) = 8.56, p = .014$, but not for the FL group, $\chi^2(2, N = 313) = 0.40, p = .82$, suggesting that those using a FL context are relatively insensitive to the amount of harm associated with the utilitarian choice.

In these experiments, we observed that reducing the severity of the negative consequences of saving the larger number of people also reduces or eliminates the FLe. This suggests that the way consequences are treated when facing moral dilemmas might depend on the language of presentation and, specifically, that the use of a FL may decrease sensitivity to negative consequences.

Analyses across experiments. In this section, we consider the entire set of experiments by performing meta-analyses and exploring the impact of FL proficiency.

We performed meta-analyses to assess the pervasiveness, robustness, and heterogeneity of the effect of language on each type of dilemma (switch and footbridge), which allowed us to reach several conclusions. First, there is a robust effect of language when the individual is faced with footbridge-type dilemmas: Choosing to save the larger number of people is more likely when using the FL than when using the NL. For an overview of the frequency of these choices made for each version of footbridge see Figure 1. Second, there is no significant effect of language when the individual is faced with switch-type dilemmas. There might be either no effect or perhaps the effect is so small that our data does not have enough power to detect it, although this option seems unlikely given the results of previous studies. Thus, it seems that there is only an effect of language when the choice to save the larger number is counterintuitive due to the aversion elicited by the dilemma's content.¹⁰ Third, the effects are very stable for both types of dilemmas.

For the details of the meta-analyses see the Appendix. For an overview of the effect sizes for each dilemma in each experiment, see Table 3; for an overview of the percentage of utilitarian choices made for each dilemma and experiment by language, see Table 4.

Another issue that we explored was the effect of FL proficiency on the magnitude of the FLe. As we have seen, the effect of proficiency was disparate for individual experiments. However, when all experiments were combined, choices to sacrifice one to save five seem to increase numerically with a decrease in proficiency.¹¹ We split the aggregate FL sample into three bins according to their self-reported proficiency (using the same criteria used for the analyses for each experiment). Those in the low-proficiency group had an average understanding of 71% and an average score

of 4.2/7 in the four abilities, those in the intermediate-proficiency group had an average understanding of 94% and an average score of 4.7/7 in the four abilities, those in the high-proficiency group all reported understanding of 100% of the dilemmas and had an average score of 5.8/7 in the four abilities.

For the footbridge dilemma, the rate of utilitarian choices significantly differed between the low- and intermediate-proficiency groups (52% vs. 42%), $\chi^2 = (1, N = 592) = 6.52, p = .01$, and the low- and high-proficiency groups (52% vs. 43%), $\chi^2 = (1, N = 591) = 5.14, p = .023$, but not between the intermediate- and high-proficiency groups (42% vs. 43%), $\chi^2(1, N = 591) = 0.08, p = .78$. Importantly, all three proficiency groups made significantly more choices to save the larger number of people (low proficiency: $\chi^2[1, N = 1,205] = 39.32, p < .001$; intermediate proficiency: $\chi^2[1, N = 1,205] = 9.46, p = .002$; high proficiency: $\chi^2[1, N = 1,204] = 11.76, p < .001$) when compared with those who used the NL (32% utilitarian choices). There were no effects of proficiency or language on choices in the switch dilemma.¹² Hence, these results replicate previous observations (Costa et al., 2014a; Geipel et al., 2015) in that the higher the proficiency in the FL, the smaller the increase in choices to save the larger number of people relative to the NL. What is fundamental, however, is that those participants with a high level of proficiency in the FL still made significantly more choices to save the larger number of people than those using their NL.

It is possible that the frequency of utilitarian choices increases as an inverse function of proficiency. To further assess the potential contribution of proficiency to the effect of language in footbridge, we also conducted a logistic regression, treating average self-reported proficiency in the four abilities as a continuous variable to evaluate if this significantly predicts choice. Self-reported proficiency was not a significant predictor of choice, $\chi^2(1) = 1.86, p = .17$. Although the Exp(B) value suggests that every decreased unit of proficiency increases the odds ratio of making the utilitarian choice by .92, this effect was not significant ($\beta = -0.08, SE = .06, p = .17$). As such, the results suggest that FL proficiency, when treated as a continuous variable, did not affect responses to the footbridge dilemma. Although this suggests

⁹ Again, we also analyzed the results for the switch dilemma. We did not find any significant effects (experiment: Wald statistic = 3.014, $p = .083$; language: Wald statistic = 0.023, $p = .88$; interaction: Wald statistic = 1.78, $p = .18$). This was expected given that the switch dilemma was identical across the three experiments.

¹⁰ It is also possible that there is a ceiling effect for switch given that roughly 80% of those using the NL routinely endorses utilitarian action, which would not leave room for an effect of language. However, this seems unlikely given that we did find the phenomenon for switch in Experiment 1a and because in at least one experiment where there would be room for an effect of language (e.g., Experiment 3e) none was found.

¹¹ Experiment 2a was excluded from these analyses because it differed from the other experiments in that (a) the order of the dilemmas was not counterbalanced across participants and (b) participants switched languages between dilemmas.

¹² There were no significant effects for the switch dilemma. Neither for proficiency: low compared with intermediate proficiency (77% vs. 78%), $\chi^2(1, N = 592) = 0.01, p = .92$, low compared with high proficiency (77% vs. 77%), $\chi^2(1, N = 591) = 0.01, p = .91$, and intermediate compared with high proficiency (78% vs. 77%), $\chi^2(1, N = 591) = 0.05, p = .83$. Nor for language (native group: 76% utilitarian): low proficiency, $\chi^2(1, N = 1205) = 0.35, p = .56$, intermediate proficiency, $\chi^2(1, N = 1205) = 0.5, p = .48$, and high proficiency, $\chi^2(1, N = 1204) = 0.19, p = .66$.

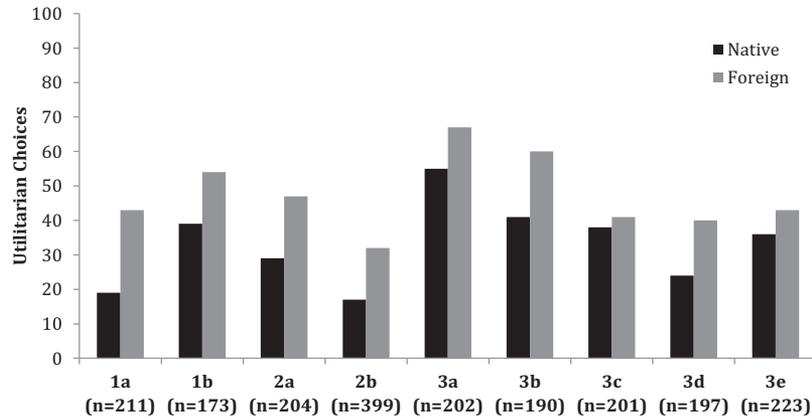


Figure 1. Percentage of utilitarian choices for footbridge-type dilemmas in the native and foreign language conditions across all experiments.

that proficiency may not play a significant role in the effect (unlike some previous analyses), this is congruent with prior results in that it argues against the effect being solely attributable to a lack of understanding.

General Discussion

In this article, we reported nine experiments (including 2,000 participants) exploring the FLE on moral choice. Two main issues were assessed: (a) the robustness of the FLE in several contexts and (b) the effect of various factors on the presence of the FLE. This has helped to better describe the phenomenon and investigate potential explanations. Next, we briefly summarize the main results of the different experiments and, subsequently, we discuss their implications.

In the first set of studies (Experiments 1a and 1b), we asked the following two questions: (a) Is the FLE robust and can the pattern

be extended to other dilemmas? and (b) Does FL vocabulary knowledge affect the magnitude of the phenomenon? The answer to these two questions is positive. The FLE was present in the footbridge dilemma and extended to the conceptually similar terrorist dilemma. On the other hand, the phenomenon was not present for the hospital dilemma, as with the switch dilemma. Furthermore, the magnitude of the effect was larger for those speakers with less knowledge of the FL, although present overall.

In the second set of studies (Experiments 2a and 2b), we asked the following two questions: (a) Is the FLE the result of increasing cognitive control due to language switching? and (b) Is the FLE a consequence of social inferences prompted by language context? The answer to these two questions is negative. The FLE was present when all participants (including those using the NL) were required to switch languages. Furthermore, the magnitude of the effect was similar irrespective of whether all participants switched

Table 3
Effect Size Indices Across all Experiments: Risk Ratio (RR) Values and the 95% Confidence Intervals for Each Experimental Condition

Type of dilemma	Experimental condition	RR	95% CI	
Switch	Exp 1a/Switch	1.168	[1.337–1.021]	
	Exp 1b/Hospital	1.128	[1.323–.962]	
	Exp 2a/Switch (language switch)	1.084	[1.203–.978]	
	Exp 2b/Switch (in-group)	.960	[1.138–.809]	
	Exp 2b/Switch (out-group)	1.035	[1.192–.899]	
	Exp 3a/Switch	.968	[1.127–.831]	
	Exp 3b/Switch (consequence)	.992	[1.127–.872]	
	Exp 3c/Switch (consequence action)	.960	[1.117–.826]	
	Exp 3d/Switch	.950	[1.106–.817]	
	Exp 3e/Switch	1.045	[1.264–.864]	
	Footbridge	Exp 1a/Footbridge	2.278	[3.574–1.452]
		Exp 1b/Terrorist	1.389	[1.925–1.002]
		Exp 2a/Footbridge (language switch)	1.623	[2.351–1.121]
Exp 2b/Footbridge (in-group)		1.851	[3.308–1.035]	
Exp 2b/Footbridge (out-group)		1.868	[2.989–1.168]	
Exp 3a/Footbridge button		1.212	[1.517–.969]	
Exp 3b/Footbridge (consequence)		1.460	[1.950–1.093]	
Exp 3c/Footbridge (consequence action)		1.078	[1.520–.765]	
Exp 3d/Footbridge disabled		1.653	[2.541–1.075]	
Exp 3e/Footbridge injured		1.209	[1.674–.873]	

Table 4
Percentage of Utilitarian Choices for Both Language Conditions and Dilemmas for Each Experiment and the Difference in this Percentage Between Language Groups for Each Dilemma and Experiment

Experiment	N	Switch		Footbridge		Difference in utilitarian choices Foreign—Native	
		Native	Foreign	Native	Foreign	Switch	Footbridge
1a	211	74	87	19	43	13	24
1b	173	73	83	39	54	10	15
2a	204	91	84	29	47	-7	18
2b	399	76	76	17	32	0	15
3a	202	78	75	55	67	-3	12
3b	190	84	83	41	60	-1	19
3c	201	79	76	38	41	-3	3
3d	197	79	75	24	40	-4	16
3e	223	64	67	36	43	3	7

languages or not. The FLe was also present when the dilemmas did not allow for social inferences about the actors. Indeed, the effect was present when the dilemmas specified the actors as either in-group or out-group members. Moreover, the magnitude of the effect did not differ between these two contexts.

In the third set of studies (Experiments 3a through 3e), we asked several questions. First, is the FLe a result of reduced action aversion? The answer to this question is unclear. The effect was marginally significant (Experiment 3a) when the action required to save the larger number of people was arguably less aversive (pushing a button) than in the original study (pushing a man). Furthermore, the effect was also present (Experiment 3b) when the choice in the dilemma was framed in terms of the consequences of inaction (letting five people die), which would presumably reduce the potential impact of action aversion. The second question we asked was whether the FLe would persist when the trade-off between the means and the ends is made highly salient. The answer to this question is no. This condition stressed both the action and the consequences: “Would you let five people die by not pushing the man?” In this case, the FLe disappeared.

Finally, the last two experiments addressed the following question: Is the FLe a result of reduced sensitivity to the negative consequences of action? The answer to this question is positive. This manipulation increased utilitarian choices for those using their NL but not those using their FL. Thus, when the negative consequences associated with the utilitarian choice were reduced (injury or disability) compared with the original study (death), the FLe was reduced or absent.

What Are All These Results Telling Us About the Nature of the FLe?

One of the most remarkable observations of our studies is the rather robust effect of language on responses to moral dilemmas. As can be appreciated in Figure 1, choices to save the larger number of people were always more frequent, in numerical terms, in the FL context than in the native one. The meta-analyses confirmed the stability of the effect across different versions of the

moral dilemmas. Hence, the phenomenon is robust and generalizable.

We also showed that some potential explanations related to the collateral effects of using a FL cannot account for the presence of the phenomenon. First, the effect is not due to an increase in cognitive control associated with language switching. Second, the effect is not due to potential social inferences about the status of the actors in the dilemmas that could be prompted by using one language or another. Furthermore, the in-group or out-group status of the actors does not seem to affect the presence of the FLe. Third, the FLe cannot be wholly attributed to random responding as a result of having a poor understanding of the dilemma (see also Geipel et al., 2015, for convergent evidence). The effect was present for participants with high vocabulary knowledge and proficiency in the FL, albeit in a smaller magnitude. In addition, subsequent analyses showed that self-reported FL proficiency was not a significant predictor of utilitarian choice, which suggests that even though proficiency does contribute to the effect, it does not account for a large portion of it. Furthermore, it seems unlikely that the effect would be due to a lack of comprehension driving random responding because this would also predict an effect of language for switch (in the opposite direction). Given that this is not the case, it is highly improbable that the use of a FL would affect the comprehension of footbridge but not switch. Having shown the pervasiveness of the phenomenon and excluded some potential explanations, we devised five studies to assess the role of actions and consequences in the presence of the FLe.

We argued that the FLe could in part be explained by reduced aversion to the action associated with saving the larger number of people. This hypothesis was based on two facts. First, people tend to be aversive to typically violent actions regardless of whether they bring harmful consequences or not. For example, people are highly aroused by pulling the trigger of a toy gun aimed at the experimenter’s head even when this has no consequences (Cushman et al., 2012). Second, it has been argued that FL contexts reduce the emotional reaction conveyed by the message (see introductory paragraphs). That is, despite understanding the message, it does not elicit as strong of an emotional reaction as in the NL. Although FL contexts may elicit a reduced emotional response compared with NL ones, and reduced action aversion would lead to an increase in choosing to save the larger number of people, the results of Experiments 3a and 3b suggest that such a reduction in aversion cannot fully account for the presence of the FLe. This is because the effect was marginally present when the action associated with the utilitarian choice arguably elicited less aversion (pushing a button instead of pushing a man). Furthermore, the effect remained when the choice in the dilemma was framed in terms of the consequences of inaction. However, we cannot conclude that a reduction in action aversion does not contribute to the phenomenon at all: The magnitude of the effect was numerically smaller in the contexts that would arguably dampen the potential contribution of action aversion (pushing a button and stressing the consequences of inaction). Moreover, the effect of varying the aversion elicited by the action was smaller, numerically speaking, on those using the FL than those using the NL. Hence, together these results suggest that although a reduction in action aversion may contribute to the magnitude of the FLe, it cannot account for the lion’s share.

The final experimental effort was devoted to assessing whether the FLe stems from a different way of evaluating the consequences of making the utilitarian choice. This was motivated by two related hypotheses. First, part of the reluctance to opt for the utilitarian choice in the footbridge dilemma could be the emotional reaction elicited by the negative consequences (being responsible for sacrificing a man's life) of making such a choice. If this reaction is attenuated when using a FL, this would then facilitate making the utilitarian choice. Second, it may be that the use of a FL prompts the dilemma to be represented in more general terms (harming one to save five), which would then decrease the potential aversion elicited by the personal repercussions of making the utilitarian choice and, thus, result in an increase of them. This issue was assessed by reducing the negative consequences associated with the choice to save the larger number of people. Indeed, this manipulation seems to have affected the magnitude of the FLe; it decreased with the severity of the negative consequences (death, disability, injury). In fact, this reduction affected choices in the NL context to a larger extent than in the foreign. These observations suggest that, at least part of the FLe may arise due to the way consequences are construed.

This pattern of results may suggest that a FL context alters the manner in which the dilemma is considered; specifically, FL use may prompt an increase in psychological distance (relative to that of NL use). If this is the case, then it would affect the way actions and consequences are valued, such that they are appraised differently than in a NL context. More concretely, this would dampen the aversion experienced when confronted with typically violent actions, lead to a reduced sensitivity to the negative consequences incurred by the choice to save the larger number of people, and increase sensitivity to the positive consequences associated with said choice (see following paragraphs). Each of these, in turn, would lead to an increase in willingness to make the trade-off of one life for five.

When the decision-making context elicits a distant perspective, this appears to reduce sensitivity to negative emotions. For example, when making decisions for others, people are less loss and risk averse than when making decisions for themselves (Beisswanger, Stone, Hupp, & Allgaier, 2003; Polman, 2012a). This could be partly due to regulatory focus (Higgins, Shah, & Friedman, 1997), which could focus people on different aspects of the decision depending on their motivation. This, in turn, could be mediated by psychological distance: People who decide for themselves are more prevention-focused and those who decide for others are more promotion-focused. This means that low psychological distance leads to focusing on the negative whereas high psychological distance leads to focusing on the positive (Beisswanger et al., 2003; Polman, 2012b). This is consistent with our finding that when using a FL, people appear to be less likely to make decisions based on avoidance of negative outcomes (harming one person) and more likely to focus on potential gains (saving lives). As argued in the preceding text, this is also congruent with the evidence regarding the FLe on economic decisions (Costa et al., 2014b; Keysar et al., 2012).

Finally, we cannot claim that the use of a FL makes one more utilitarian per se. The current methods do not allow for the dissociation of deontological and utilitarian inclinations and

further research should apply a process dissociation approach (e.g., Conway & Gawronski, 2013). Regardless, what we can claim given the current set of results is that there is an effect of language on choice. As we have argued, this may be due to the use of a FL prompting different considerations regarding the means and consequences.

It is critical to understand how and when our moral choices may be affected by language. The nine studies reported here contribute to our understanding of the factors and processes that affect moral choice by providing support for the pervasiveness of the FLe across different contexts. Furthermore, the current work offers initial insights on the potential processes underlying the impact of a FL on moral choice. Although most of us might never face a choice as extreme as the one made by Alexander Holmes and the crew as the lifeboat began to sink, we are all faced with moral dilemmas—whether in the workplace, in our political or religious stances, or in our personal lives. The fact that these important decisions could be influenced by the “nativeness” of the language we use has far-reaching implications given the millions of people who use a FL every day.

References

- Aguilar, P., Brussino, S., & Fernández-Dols, J. M. (2013). Psychological distance increases uncompromising consequentialism. *Journal of Experimental Social Psychology, 49*, 449–452. <http://dx.doi.org/10.1016/j.jesp.2013.01.002>
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist, 63*, 602–614. <http://dx.doi.org/10.1037/0003-066X.63.7.602>
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*, 154–161. <http://dx.doi.org/10.1016/j.cognition.2011.05.010>
- Beisswanger, A. H., Stone, E. R., Hupp, J. M., & Allgaier, L. (2003). Risk taking in relationships: Differences in deciding for oneself versus for a friend. *Basic and Applied Social Psychology, 25*, 121–135. http://dx.doi.org/10.1207/S15324834BASP2502_3
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley. <http://dx.doi.org/10.1002/9780470743386>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111. <http://dx.doi.org/10.1002/jrsm.12>
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185–216. <http://dx.doi.org/10.1177/135910457000100301>
- Broeders, R., Van Den Bos, K., Müller, P. A., & Ham, J. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology, 47*, 923–934. <http://dx.doi.org/10.1016/j.jesp.2011.03.018>
- Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making, 6*, 580–592.
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2*, 84–92. <http://dx.doi.org/10.1093/scan/nsm001>
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology, 29*, 23–40.
- Colbeck, K. L., & Bowers, J. S. (2012). Blinded by taboo words in L1 but not L2. *Emotion, 12*, 217–222. <http://dx.doi.org/10.1037/a0026387>

- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*, 216–235. <http://dx.doi.org/10.1037/a0031021>
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.
- Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apestequia, J. (2014a). “Piensa” twice: On the foreign language effect in decision making. *Cognition, 130*, 236–254. <http://dx.doi.org/10.1016/j.cognition.2013.11.010>
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apestequia, J., Heafner, J., & Keysar, B. (2014b). Your morals depend on language. *PLoS One, 9*(4), e94842. <http://dx.doi.org/10.1371/journal.pone.0094842>
- Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review, 17*, 273–292. <http://dx.doi.org/10.1177/1088868313495594>
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion, 12*, 2–7. <http://dx.doi.org/10.1037/a0025071>
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*, 1082–1089. <http://dx.doi.org/10.1111/j.1467-9280.2006.01834.x>
- Dewaele, J. M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development, 25*, 204–222. <http://dx.doi.org/10.1080/01434630408666529>
- Dewaele, J. M. (2011). Self-reported use and perception of the L1 and L2 among maximally proficient bi-and multilinguals: A quantitative and qualitative investigation. *International Journal of the Sociology of Language, 208*, 25–51. <http://dx.doi.org/10.1515/ijsl.2011.011>
- Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121–127. <http://dx.doi.org/10.1016/j.cognition.2014.09.006>
- Eilola, T. M., & Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *The International Journal of Bilingualism, 15*, 353–369. <http://dx.doi.org/10.1177/1367006910379263>
- Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion, 21*, 1064–1076. <http://dx.doi.org/10.1080/02699930601054109>
- Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.; pp. 237–253). New York, NY: Russell Sage Foundation.
- Foot, P. (1978). The problem of abortion and the doctrine of the double effect. *Virtues and vices* (pp. 19–32). Oxford, UK: Blackwell.
- Foroni, F. (2015). Do we embody second language? Evidence for ‘partial’ simulation during processing of a second language. *Brain and Cognition, 99*, 8–16. <http://dx.doi.org/10.1016/j.bandc.2015.06.006>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS One, 10*(7), e0131529. <http://dx.doi.org/10.1371/journal.pone.0131529>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition, 154*, 34–39. <http://dx.doi.org/10.1016/j.cognition.2016.05.010>
- Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Update, 11*, 322–323. <http://dx.doi.org/10.1016/j.tics.2007.06.004>
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*, 364–371. <http://dx.doi.org/10.1016/j.cognition.2009.02.001>
- 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*, 389–400. <http://dx.doi.org/10.1016/j.neuron.2004.09.027>
- 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*, 2105–2108. <http://dx.doi.org/10.1126/science.1062872>
- Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied, 21*, 117–129. <http://dx.doi.org/10.1037/xap0000044>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834. <http://dx.doi.org/10.1037/0033-295X.108.4.814>
- Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25*(2–3), 223–247. <http://dx.doi.org/10.1080/01434630408666530>
- Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics, 24*, 561–579. <http://dx.doi.org/10.1017/S0142716403000286>
- Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. *Bilingual Education and Bilingualism, 56*, 257–283.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504. <http://dx.doi.org/10.1037/1082-989X.3.4.486>
- Higgins, E. T., Shah, J., & Friedman, R. (1997). Emotional responses to goal attainment: Strength of regulatory focus as moderator. *Journal of Personality and Social Psychology, 72*, 515–525. <http://dx.doi.org/10.1037/0022-3514.72.3.515>
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558. <http://dx.doi.org/10.1002/sim.1186>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal (Clinical Research Ed.), 327*, 557–560. <http://dx.doi.org/10.1136/bmj.327.7414.557>
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychological Methods, 11*, 193–206. <http://dx.doi.org/10.1037/1082-989X.11.2.193>
- 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*, 489–496. <http://dx.doi.org/10.1037/xlm0000179>
- Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience, 10*, 551–560. <http://dx.doi.org/10.1080/17470919.2015.1023400>
- Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). ‘Utilitarian’ judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193–209. <http://dx.doi.org/10.1016/j.cognition.2014.10.005>
- Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience, 7*, 393–402. <http://dx.doi.org/10.1093/scan/nsr005>

- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, *23*, 661–668. <http://dx.doi.org/10.1177/0956797611432178>
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908–911. <http://dx.doi.org/10.1038/nature05631>
- Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289. <http://dx.doi.org/10.1146/annurev.psych.58.110405.085654>
- Miller, R. M., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*, 707–718. <http://dx.doi.org/10.1111/spc3.12066>
- Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, *14*, 573–587. <http://dx.doi.org/10.1037/a0035361>
- 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*, 549–557. <http://dx.doi.org/10.1111/j.1467-9280.2008.02122.x>
- Moretto, G., Ladavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*, 1888–1899. <http://dx.doi.org/10.1162/jocn.2009.21367>
- Nakamura, K. (2013). A closer look at moral dilemmas: Latent dimensions of morality and the difference between trolley and footbridge dilemmas. *Thinking & Reasoning*, *19*, 178–204. <http://dx.doi.org/10.1080/13546783.2013.768551>
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*, 530–542. <http://dx.doi.org/10.1016/j.cognition.2005.07.005>
- Oganian, Y., Korn, C. W., & Heekeren, H. R. (2016). Language switching—but not foreign language use per se—reduces the framing effect. *Journal of Learning, Memory, and Cognition*, *42*, 140–148. <http://dx.doi.org/10.1037/xlm0000161>
- Open Doors. (2014, November 17). 2014 “fast facts.” Retrieved from <http://www.iie.org/Research-and-Publications/Open-Doors>
- Pavlenko, A. (2004). ‘Stop doing that, ia komu skazala!’: Language choice and emotions in parent–child communication. *Journal of Multilingual and Multicultural Development*, *25*, 179–203. <http://dx.doi.org/10.1080/01434630408666528>
- Pavlenko, A. (2005). *Emotions and multilingualism*. New York, NY: Cambridge University Press.
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*, 405–428. <http://dx.doi.org/10.1080/00207594.2012.743665>
- Petrinovich, L., & O’Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*, 145–171. [http://dx.doi.org/10.1016/0162-3095\(96\)00041-6](http://dx.doi.org/10.1016/0162-3095(96)00041-6)
- Polman, E. (2012a). Effects of self–other decision making on regulatory focus and choice overload. *Journal of Personality and Social Psychology*, *102*, 980–993. <http://dx.doi.org/10.1037/a0026966>
- Polman, E. (2012b). Self–other decision making and loss aversion. *Organizational Behavior and Human Decision Processes*, *119*, 141–150. <http://dx.doi.org/10.1016/j.obhdp.2012.06.005>
- Puntoni, S., de Langhe, B., & van Osselaer, S. M. (2009). Bilingualism and the emotional intensity of advertising language. *The Journal of Consumer Research*, *35*, 1012–1025. <http://dx.doi.org/10.1086/595022>
- Railton, P. (2016). Moral learning: Why learning? Why moral? And why now? *Cognition*, S0010-0277(16)30205–0. <http://dx.doi.org/10.1016/j.cognition.2016.08.015>
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295–315). New York, NY: Russell Sage Foundation.
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://R-project.org>
- Segalowitz, N., Trofimovich, P., Gattbonton, E., & Sokolovskaya, A. (2008). Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon*, *3*, 47–71. <http://dx.doi.org/10.1075/ml.3.1.05seg>
- Sheikh, N. A., & Titone, D. (2016). The embodiment of emotional words in a second language: An eye-movement study. *Cognition and Emotion*, *30*, 488–500.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*, 454–458. <http://dx.doi.org/10.1016/j.cognition.2011.01.018>
- 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*, 1077–1090. <http://dx.doi.org/10.1080/02699930601054133>
- Swann, W. B., Jr., Gómez, A., Dovidio, J. F., Hart, S., & Jetten, J. (2010). Dying and killing for one’s group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science*, *21*, 1176–1183. <http://dx.doi.org/10.1177/0956797610376656>
- Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*, 227–229.
- Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*, 1395. <http://dx.doi.org/10.2307/796133>
- Thomson, J. J. (1986). *Rights, restitution, and risk: Essays in moral theory*. Cambridge, MA: Harvard University Press.
- Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, *40*, 923–930. <http://dx.doi.org/10.1177/0146167214530436>
- Trémolière, B., & De Neys, W. (2013). Methodological concerns in moral judgement research: Severity of harm shapes moral decisions. *Journal of Cognitive Psychology*, *25*, 989–993. <http://dx.doi.org/10.1080/20445911.2013.841169>
- Trémolière, B., Neys, W. D., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*, 379–384. <http://dx.doi.org/10.1016/j.cognition.2012.05.011>
- United States v. Holmes. (1842). Circuit Court, E. D. Pennsylvania. 26 F. Cas. 360. Buffalo Criminal Law Center.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*, 476–477. <http://dx.doi.org/10.1111/j.1467-9280.2006.01731.x>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. <http://dx.doi.org/10.18637/jss.v036.i03>
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18*, 247–253. <http://dx.doi.org/10.1111/j.1467-9280.2007.01884.x>
- Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral judgment. In R. G. Morison (Ed.), *The Oxford handbook of thinking and reasoning* (pp. 364–389). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780199734689.013.0019>
- Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition*, *126*, 364–372. <http://dx.doi.org/10.1016/j.cognition.2012.11.002>
- Winkler, H. (2013). The emotional Stroop task and emotionality rating of

negative and neutral words in late Thai–English bilinguals. *International Journal of Psychology*, 48, 1090–1098. <http://dx.doi.org/10.1080/00207594.2013.793800>

Wu, Y. J., & Thierry, G. (2012). How reading in a second language protects your heart. *The Journal of Neuroscience*, 32, 6485–6489. <http://dx.doi.org/10.1523/JNEUROSCI.6119-11.2012>

Appendix

Meta-analyses

We evaluated three aspects of the data: (1) The pervasiveness of the effect of language by determining effect sizes for each type of dilemma and experiment, (2) the robustness of the effect of language across experiments for each type of dilemma, (3) the heterogeneity of the effects for each type of dilemma. The procedures used are meta-analytic techniques for synthesizing values of effect size indices (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009). We have combined the results from the switch-type and the footbridge-type dilemmas separately. To preserve the assumptions of the statistical models, we only combined results from samples of different participants.

To determine the robustness of the foreign language effect for each dilemma in each experiment, we have obtained the risk ratio index, based on the proportion of utilitarian responses given by those using the foreign and the NLs. Those proportions being p_f and p_n , the index is calculated as risk ratio = p_f/p_n (Fleiss & Berlin, 2009). When in a given experiment there is no language effect, the risk ratio value is around 1. Values significantly above 1 reflect higher rates of utilitarian responses when using the foreign than the NL. When the interval includes the value 1, the null hypothesis (no association between the language and decision) must be maintained, whereas when the interval completely exceeds the value 1 (lower bound >1), it can be concluded that there is a significant language effect in the expected direction.

As expected, in almost all of the switch-type dilemmas there is no evidence of a significant association between the language and the propensity to make utilitarian choices. Only in 1 of 10 experiments (Experiment 1a) the result is statistically significant (close to the expected 5% Type I error rate). On the contrary, in 7 of 10 of the footbridge-type dilemmas there is evidence of a significant association between language and choice. Next, we analyzed across the experiments for each type of dilemma, combining the 10 risk ratio values. For each set of the two types of dilemma, we have combined these estimates under a random effects model, weighting the risk ratio values by the inverse variance method (Hedges & Olkin, 1985). Two main models are usually employed to combine independent estimates (Borenstein, Hedges, Higgins, & Rothstein, 2010; Hedges & Vevea, 1998). Fixed effect models assume that all the studies are essentially identical; then, between-study differences are due to sampling error alone. Random effects models

assume that, in addition to sampling error there are other sources of between study variability; as a consequence, the studies combined for a random effects model estimate different parametric values. In general, in psychology the assumption that there is some variability in the parameter across the studies is more realistic. Random effects models reflect this feature adding a separate variance term (the so-called specific variance).

The specific variance has been estimated by the restricted maximum likelihood method (Raudenbush, 2009). Calculations have been carried out using METAFOR (Viechtbauer, 2010), an R package for meta-analysis (R Development Core Team, 2010).

As expected, the results reveal that we should maintain the null hypothesis for switch-type dilemmas and reject it for footbridge-type ones: The combined estimate with the switch-type dilemmas is 1.033 (95% CI: 1.083–0.986), whereas for the footbridge-type dilemmas it equals 1.437 (95% CI: 1.639–1.260). The conclusion is clear. The synthesis of 10 independent experimental conditions that employed switch-type dilemmas does not show evidence of a significant association between the language and the propensity to choose the utilitarian response. On the contrary, the synthesis of 10 experimental conditions that employed footbridge-type dilemmas shows evidence of a significant association. The direction of the effect is as expected: when faced with footbridge-type dilemmas, utilitarian choices are more probable when using the foreign language than the native one.

Finally, we assessed the degree of heterogeneity of the effect size values, as is important when doing meta-analyses. This is accomplished by the Cochrane- Q test (Hedges & Olkin, 1985), the I^2 index (Higgins, Thompson, Deeks, & Altman, 2003; Higgins & Thompson, 2002), and the value of the specific variance estimated. The Q test yields no significant effects in either set of experimental conditions [switch-type dilemmas: $Q(9) = 9.12$, $p = .43$; footbridge-type dilemmas: $Q(9) = 12.68$, $p = .18$]. However, this test has low power when the sample of estimates is small, as in our study ($k = 10$). The I^2 index is a descriptive value that reflects the percentage in which the observed variability exceeds the expected variability due to sampling. It is often recommended as a complement of the Q test (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). The 25%, 50% and 75% values of I^2 are taken as the references for low, medium, and high degrees of

(Appendix continues)

variability (Higgins et al., 2003). In our results the values are $I^2 = 6.1\%$ for the switch-type dilemmas and $I^2 = 26.9\%$ for the footbridge-type dilemmas.

The empirical variance of the values must be compared with the degree of variability expected by mere sampling. When the values show a significantly higher level of heterogeneity, it is explained by the random effect component of the model. This reflects the fact that the studies are performed under a variety of circumstances, participants, procedures, and so forth; these factors yield some variability in the actual effects that the individual studies are estimating. So, fitting a random effects model involves estimating the specific variance of the random factor. As the switch-type dilemmas are in all but one of our experiments (Experiment 1b) essentially the same (the standard switch dilemma), not very much heterogeneity is expected. On the contrary, the footbridge-type dilemmas employed along the experiments have been presented with some variations in the linguistic context, in the text concern-

ing the consequences for the utilitarian sacrifice (death, disability, injury) or the details of utilitarian action, or highlighting some aspects of the decisional context. This should generate some extra variability in the effects that would be reflected in the specific variance. As expected, the variance is higher for the footbridge-type than for the switch-type dilemmas. When combining the estimates for each type of dilemma we find that the specific variances equal 0.0003 and 0.0118 for the switch- and footbridge-type dilemmas, respectively. The pattern of results of the Q -test, the I^2 index and the specific variance values indicates that there is more heterogeneity in the footbridge-type than in the switch-type dilemmas.

Received February 1, 2016

Revision received September 27, 2016

Accepted October 1, 2016 ■