



The foreign language effect on decision-making: A meta-analysis

Riccardo Circi¹ · Daniele Gatti² · Vincenzo Russo¹ · Tomaso Vecchi^{2,3}

Accepted: 20 December 2020
© The Psychonomic Society, Inc. 2021

Abstract

In recent years, a growing body of literature has shown that being in a foreign language (FL) context affects the way in which people make choices. This phenomenon is known as the foreign language effect (FLE). The FLE affects both moral decision-making and risk-aversion tendencies, but no cumulative evidence is available. Herein, we aimed to estimate, through a meta-analytical approach, the effect of being in an FL context as compared with that of a native language (NL). We found 17 studies matching our criteria and, in total, 47 experiments were included ($N = 38$ investigated the FLE in the moral decision-making domain; $N = 9$ investigated the FLE in the risk-aversion domain). Results showed that FL affects participants' decisions as compared with NL in both the moral decision-making and risk-aversion domains, inducing participants to be more willing to accept harm in order to maximize outcomes in the former and reducing risk aversion in the latter. In addition, two metaregressions were performed on the studies that investigated the moral decision-making domain in order to assess whether participants' proficiency in the FL, or NL–FL similarity, moderated the observed effect. Our findings indicate that proficiency in the FL does not moderate the observed effect, while NL–FL similarity does. Our results support previous findings on the FLE and provide suggestions for future research.

Keywords Foreign language effect · FLE · Decision-making · Moral dilemma

Abbreviations

FLE Foreign language effect
FL Foreign language
NL Native language

Introduction

In recent years, a growing body of literature has shown that being in a foreign language (FL) context affects people's

decision-making (for a review, see Hadjichristidis, Geipel, & Keysar, 2019). This phenomenon was first described by Keysar, Hayakawa, and An (2012) and is known as the foreign language effect (FLE). The authors found that a group of participants who answered the Asian disease dilemma in an FL exhibited less framing effect than those who answered the dilemma using their native language (NL). The Asian disease problem, first proposed by Tversky and Kahneman (1981), describes an epidemic which, without treatment, would have severe consequences (e.g., death of 600 people). Generally, in the Asian disease problem, participants have to choose between a safe option and a risky option to deal with an epidemic. Interestingly, Tversky and Kahneman (1981) found that participants' judgements are influenced by how the two options are presented. That is, options could be presented in a gain frame (e.g., "If you choose Option A, 200 people will be saved; if you choose Option B, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved") or in a loss frame (e.g., "If you choose Option A, 400 people will die; if you choose Option B, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die"). In the gain frame condition, participants tend to choose the safe option (i.e., Option A); on the contrary, in the loss frame condition, participants tend to choose the risky

Riccardo Circi and Daniele Gatti contributed equally to this work.

- ✉ Riccardo Circi
riccardo.circi@studenti.iulm.it
- ✉ Daniele Gatti
daniele.gatti@unipv.it

- ¹ Behavior and Brain Lab IULM–Neuromarketing Research Center, Università IULM, Via Carlo Bo 1, 20143 Milan, Italy
- ² Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy
- ³ Cognitive Psychology Unit, IRCCS Mondino Foundation, Pavia, Italy

option (i.e., Option B; Tversky & Kahneman, 1981). It is important to point out that expected values of Options A and B are the same in the gain frame and in the loss frame. Therefore, the different pattern of responses in the two framing is considered a violation of the rules of rational choice, according to which the choices should be based only on the expected values and not on the framing in which options are described (Von Neumann & Morgenstern, 1947). Critically, manipulating language presentation, Keysar et al. (2012) found that when the Asian disease problem is presented in an FL, the framing effect is reduced, and in both framing participants tend to choose the safe option in similar extent.

Several other studies have investigated the FLE, both by replicating the phenomenon on the framing effect (Costa, Foucart, Armon, Aparici, & Apesteguia, 2014a; Winkler, Ratitamkul, Brambley, Nagarachinda, & Tiencharoen, 2016) or by addressing the impact of the use of an FL on other aspects of decision-making: It has been shown that using an FL reduces risk aversion (Costa, Foucart, Armon, et al., 2014a; Hayakawa, Tannenbaum, Costa, Corey, & Keysar, 2017b), longevity of the repetition truth bias (Nadarevic, Plier, Thielmann, & Darancó, 2018), self-serving bias (Van Hugten & van Witteloostuijn, 2018), causality bias (Díaz-Lago & Matute, 2018), and the hot-hand fallacy (Gao, Zika, Rogers, & Thierry, 2015). Similarly, participants who use an FL have a lesser tendency to lie and dishonestly inflate their earnings as compared with participants who use their NL (Bereby-Meyer et al., 2018). Furthermore, reading the descriptions of disgusting but healthy and sustainable products (e.g., biscuits made with worm meal) in an FL increases the declared willingness to consume them (Geipel, Hadjichristidis, & Klesse, 2018).

In addition, many studies on the FLE have addressed how being in an FL context can affect decision-making when there is the possibility to accept harm in order to maximize outcomes. It has been shown that presenting participants with an FL version of the footbridge dilemma leads participants to be more willing to sacrifice one life in order to save five in a hypothetical scenario (Costa, Foucart, Hayakawa, et al., 2014b). Many other studies have shown that people are typically more willing to accept harm in order to maximize outcomes when encountering dilemmas written or spoken in an FL (Brouwer, 2019; Cicolletti, McFarlane, & Weissglass, 2016; Costa, Foucart, Hayakawa, et al., 2014b; Geipel, Hadjichristidis, & Surian, 2015a, 2015b, 2016).

The growing interest in literature on the FLE has led to the formulation of several explanatory hypotheses. In the following sections we discuss some of the main ones.

The role of cognitive control

Although robust evidence has been reported for the FLE, the underlying cognitive mechanisms are not yet clear. It has been

proposed that two systems are involved in decision-making processes—a fast, intuitive, automatic, and affective system (System 1), and another which is deliberate and rational but also more effortful (System 2; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Kahneman, 2003; Stanovich & West, 2000). Within this perspective, one of the possible explanations of the FLE is related to a reduced reliance on System 1 and/or an increased reliance on System 2. Indeed, Kahneman (2011) argued that every element that increases cognitive tension, such as processing dilemmas in an FL, could favor the processes of System 2, reducing the impact of System 1. Thus, the reduced tendency to be influenced by cognitive biases when reading dilemmas in an FL would be explained by an enhanced cognitive control. However, other studies have shown that being in an FL context does not reduce cognitive biases when participants are presented with emotional neutral tasks (e.g., the cognitive reflection test; the Moses illusion task; Costa, Foucart, Armon, et al., 2014a; Geipel et al., 2015a; Mækelæ & Pfuhl, 2019), and can make people more prone to outcome bias (Geipel et al., 2016; but cf. Costa et al., 2018, which found no evidence of increased outcome bias in moral judgments using an FL). Within the moral decision-making domain, the FLE could be interpreted as a greater preference for utilitarian options (driven by controlled cognitive processes) over deontological options (driven by automatic emotional responses) when moral dilemmas are presented in an FL. Nevertheless, through employing a process dissociation approach (Conway & Gawronski, 2013; Jacoby, 1991) to independently assess deontological and utilitarian response inclinations, it has been shown that reading dilemmas in an FL reduces deontological tendencies without enhancing utilitarian tendencies (Hayakawa, Tannenbaum, et al., 2017), or reduces both deontological and utilitarian inclinations (Muda, Niszczoła, Bialek, & Conway, 2018).

It has also been shown that language switching can enhance cognitive control (Wu & Thierry, 2013), leading to the hypothesis that the FLE could be due to the language switching required in order to comprehend instructions in one language and perform the task in another (Oganian, Korn, & Heekeren, 2016). However, other studies have failed to replicate this finding (Corey et al., 2017; Driver, 2020).

The role of emotions

Another possible explanation is that the FLE is caused by a dampened emotionality of scenarios when these are presented in an FL. This perspective is consistent with a large body of literature, according to which an FL is perceived as less emotional than an NL (for a review, see Caldwell-Harris, 2015; Pavlenko, 2012, 2017). While the NL is acquired in an emotionally rich and naturalistic context, an FL is often acquired in a classroom context, therefore, in an emotionally neutral context and beyond that critical period within which the emotions

are associated with the morphological, orthographic, and semantic aspects of words. Similarly, it has been proposed that the reduced emotionality of the FL might concern words with a negative value, as conversations in the FL are generally centered on emotionally positive exchanges, leading to greater probability of being incorporated during the acquisition of an FL (Boucher & Osgood, 1969; for recent experimental evidence, see Foroni, 2015; Jończyk, Korolczuk, Balatsou, & Thierry, 2019; Sheikh & Titone, 2016; Wu & Thierry, 2012). Thus, frequency of use, acquisition order, linguistic dominance, context of acquisition, age of acquisition, and proficiency are considered as key aspects affecting the perceived emotionality of language (Boucher & Osgood, 1969; Dewaele, 2004; Pavlenko, 2004, 2012). Consistent with this theory, it has been shown that moral judgments are harsher in people with a higher FL proficiency (Geipel et al., 2015a) and early bilinguals show no FLE (Brouwer, 2019; Čavar & Tytus, 2017; Wong & Ng, 2018), probably because higher linguistic competence and earlier age of acquisition increase the emotional impact of the FL, making it closer to that of the NL. Nevertheless, the absence of the FLE in early bilinguals is debated (Białek & Fugelsang, 2019; Brouwer, 2020). The modality in which dilemmas are presented (written or auditory) can influence the extent to which the FLE occurs in highly proficient bilinguals (Brouwer, 2019, 2020).

The role of social norms and linguistic similarity

It has been proposed that the FLE in the moral decision-making domain could be due to a reduced activation of social and moral norms and not only by an attenuation of emotions, since dampened emotionality related to the use of an FL also occurs in dilemmas not affected by the FLE (i.e., trolley dilemma; Geipel et al., 2015b). That is, the use of an FL could limit the access to relevant autobiographical memories (including normative knowledge) that keep traces of the linguistic context in which they were learned (i.e., NL). Thus, participants could be more willing to violate a social norm (e.g., “Do not cause harm to anybody”) in order to maximize outcomes (e.g., causing a one-person death in order to save five people). This explanation is supported by recent work that demonstrated that sensitivity to norms is reduced while reading moral dilemmas in an FL (Białek, Paruzel-Czachura, & Gawronski, 2019).

Recent studies have also shown that NL–FL similarity could be another relevant factor in determining the FLE (Dylman & Champoux-Larsson, 2020; Miozzo et al., 2020): When the FL has a strong cultural influence (e.g., the English language in Sweden) or the NL and FL are linguistically similar (e.g., the Norwegian language and the Swedish language), the FLE does not occur, despite a significant difference in proficiency between NL and FL.

Our study

Many other studies are required to shed light on the FLE and underlying mechanisms, given the implications that this phenomenon could have in all those international contexts in which crucial decisions are made using an FL. However, any discussion on the implications of the FLE should also take the effect size into consideration. To the best of our knowledge, there are no published meta-analyses on the FLE that has addressed issues like effect size, heterogeneity, and publication bias. Herein, we cover this lack in the literature, giving both a contribution to future research and insight on the implication of the FLE outside of the laboratory environment.

Material and methods

Study selection and inclusion criteria

To identify potential studies for inclusion in the meta-analysis, we first conducted a systematic literature search using the scientific online database Scopus. The time window of our literature search was from January 2012 to December 2020. Search criteria were set as: ((FLE) OR (foreign language effect)) AND ((moral dilemma) OR (decision making)).

The inclusion criteria adopted were (a) between-subject study design; (b) direct comparison between the NL group and FL group¹; (c) sample composed by bilinguals living in their L1 country; (d) proficiency level explicitly reported; (e) absence of switching between instruction and materials language; (f) experiments that compared risk aversion or moral decision-making in the NL group and FL group. It should be noted that in the moral decision-making domain we included personal dilemmas that assessed the willingness to violate a social norm in order to maximize utility.

The literature search identified 125 articles. After full-text examination, we found 17 studies matching our criteria. If a study included more than one experiment, in our data set that study had as many rows as experiments. The effect size was calculated separately for each experiment. In total, 47 experiments were included (see Table 1).

The studies included in our meta-analysis were conducted between 2012 and 2020. The sample size varied from 18 to 800. Of the 47 experiments included, 38 investigated the FLE in the moral decision-making domain and nine in the risk-

¹ The framing effect in the Asian disease problem (e.g., Tversky & Kahneman, 1981) is often inferred by different patterns of answers to gain versus loss frames. However, the direct comparison between NL and FL is almost never reported, and conclusions on the FLE are drawn based on the lack of significance in the FL condition (e.g., Keysar et al., 2012, first experiment). Thus, the experiments employing the Asian disease problem were excluded because, in these cases, there is no general consensus on how to quantify the effect sizes (cf. Oganian et al., 2016).

Table 1 Studies included in meta-analyses. For each experiment (Exp) Hedges' *g*, variance, linguistic conditions, numerosity and domain are reported

Study	Exp	Hedges's <i>g</i>	Var	FL	NL	<i>N</i>	Proficiency (%)	Similarity	Domain
Costa, Foucart, Hayakawa, et al., 2014b	1	0.3	0.04	Spanish	English	112	57	Different	Moral decision-making
	2	0.26	0.05	English	Korean	80	57	Different	
	3	0.22	0.04	French	English	107	57	Different	
	4	0.95	0.27	Hebrew	English / Spanish	18	57	Different	
Geipel et al., 2015b	5	0.52	0.01	Spanish	English	328	77.92	Different	
	6	0.64	0.01	English	Spanish	397	80.83	Different	
	7	0.63	0.04	English / German	Italian	105	71.20	Different	
	8	0.33	0.03	English	Chinese	152	63	Different	
Chan, Gu, Ng, & Tse, 2016	9	0.4	0.06	English	German	72	81	Similar	
	10	0.5	0.03	English	Chinese	144	65.60	Different	
	11	0.54	0.02	English	Spanish	211	72.14	Different	
	12	0.3	0.02	English	Spanish	173	67.14	Different	
Corey et al., 2017	13	0.34	0.01	English	Spanish	399	72.50	Different	
	14	0.24	0.02	English	Spanish	202	64.64	Different	
	15	0.38	0.02	English	Spanish	190	66.79	Different	
	16	0.06	0.02	English	Spanish	201	72.50	Different	
Hayakawa, Lau, Holtzmann, Costa, & Keysar, 2017a	17	0.15	0.02	English	Spanish	223	73.93	Different	
	18	0.34	0.02	English	Spanish	197	67.14	Different	
	19	-0.21	0.02	English	German	214	71	Similar	
	20	0.12	0.02	Spanish	English	242	75.57	Different	
Shin & Kim, 2017	21	0.2	0.02	English	Spanish	195	77.14	Similar	
	22	-0.04	0.02	English	German	211	69.14	Similar	
	23	-0.24	0.02	English	German	209	74.14	Similar	
	24	0.08	0.02	German	English	161	73.86	Similar	
Hayakawa & Keysar, 2018	25	0.62	0.02	English	Korean	161	43.80	Different	
	26	0.14	0.01	English	German	800	71.43	Similar	
	27	-0.2	0.03	English	Polish	140	69.20	Different	
	28	-0.08	0.02	English	Polish	204	81.90	Different	
Bialek & Fugelsang, 2019	29	-0.11	0.03	German	Polish	138	71.70	Different	
	30	0.12	0.03	Spanish	Polish	163	55.40	Different	
	31	0.29	0.03	French	Polish	129	64.40	Different	
	32	0.06	0.07	English	Dutch	60	74	Similar	
Brouwer, 2019	33	0.45	0.07	English	Dutch	60	71.40	Similar	
	34	0.28	0.03	English	Dutch	154	80.40	Similar	
	35	0.13	0.02	English	Swedish	198	76.20	Similar	
	36	0.45	0.02	French	Swedish	175	46.20	Different	
Dylman & Champoux- Larsson, 2020	37	0.07	0.01	Norwegian	Swedish	305	51.40	Similar	
	38	-0.06	0.01	Swedish	Norwegian	295	62.20	Similar	
	39	0.44	0.03	English	Korean	146	44		
	40	0.55	0.07	Spanish	English	54	63.33		
Keysar et al., 2012	41	0.23	0.01	English	Spanish	300	68.57		Risk aversion
	42	0.3	0.02	English	Spanish	245	71.43		
	43	0.42	0.04	English	Italian	92	82.00		
	44	0.46	0.03	English	Italian	123	80		
Costa, Foucart, Amon, et al., 2014a	45	0.05	0.1	English	Thai	40	61.43		
	46	0.15	0.03	English	Chinese	155	52.86		
	47	0.07	0.04	Spanish	English	97	68.57		

Note. For each experiment, Hedges's *g*, variance, linguistic conditions, numerosity, and domain are reported.

aversion domain. Participants' NL was Spanish in 12 experiments, English in seven, German in five, Polish in five, Italian in three, Korean in three, Chinese in three, Dutch in three, Swedish in three, Thai in one, Norwegian in one, and Spanish/English with mixed sample in one. By contrast, the most frequently used FL was English (in 32 experiments), while the other languages were: Spanish in six, French in

three, German in two, Swedish in one, Norwegian in one, Hebrew in one, and English/German in one.

Effect size calculation

Answers related to risk aversion and outcome maximization obtained in an FL condition as compared with an NL

condition were our dependent variables of interest for “risk domain” and for “moral decision-making” meta-analyses, respectively. From each study we extracted NL, FL, sample size, participants’ proficiency in the FL, and the mean and standard deviation of participants’ performance in the various conditions for the variable of interest.

The effect size used was Hedges’s g (Borenstein, Hedges, Higgins, & Rothstein, 2009), which is a standardized mean difference that accounts for sampling variance difference between conditions. Effect size and variance calculation were performed using RStudio (RStudio Team, 2015) and its package *compute.es* (Del Re, 2013). The package *compute.es* uses as conversion formulas those reported by Cooper, Hedges, and Valentine (2009).

When more than one NL or FL per study were employed, the experiments were separated in multiple rows, (i.e., multiple NL vs. multiple FL), where possible. If, within one experiment (i.e., with the same sample), more than one effect was reported, the effect sizes were aggregated using RStudio (RStudio Team, 2015) and its package *MAd* (Del Re & Hoyt, 2014) using the function *agg*.

Data analysis

In order to assess whether performing a task in an FL modulates participants’ answers, and to quantify the effect of the modulation, we performed two separate meta-analyses with a random-effects method, one for each domain (i.e., moral decision-making or risk). We also performed two meta-regressions on the studies that investigated the moral decision-making domain in order to assess whether participants’ proficiency² in the FL or if NL–FL similarity (cf. Dylman & Champoux-Larsson, 2020; Miozzo et al., 2020) moderated the observed effect. In particular, those languages that belonged to the same linguistic branch were considered similar (i.e., English and German are both Germanic languages).

We decided not to add the FL as moderator, due to the low number of studies that employed languages different from English. We also chose not to perform a meta-regression on the studies that investigated decisions in the risk domain, due to the low numerosity of each language group.

The within studies heterogeneity was evaluated using the Q test. A significant p value of the Q test implies that the observed within-studies variance can be explained by other variables besides the FLE. In addition, we used Higgins’s I^2 as the index of heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003); this provides the percentage of the total variability in the effect size estimation that could be attributed to heterogeneity among the true effects (heterogeneity is considered high if $I^2 > 75\%$; Higgins et al., 2003).

² Since the included studies used different proficiency scales, we converted the data into percentages.

The publication bias was evaluated using funnel plots and the trim-and-fill method (Duval, 2005). The trim-and-fill method provides an estimate of the number of studies missing from the meta-analysis due to the suppression of the most extreme results on one side (generally the left; i.e., nonsignificant results) of the funnel plot. To further explore the publication bias, the Egger’s test (Egger, Davey Smith, Schneider, & Minder, 1997) was performed. The Egger’s test examines the correlation between the various effect sizes and their sampling variances (i.e., if the funnel plot is asymmetric), and a significant p value indicates publication bias (Lin & Chu, 2018). To explore the robustness of the results, a fail-safe number of studies was computed to obtain an estimate of how many null studies should be included in order to render the observed effect nonsignificant (Rosenthal, 1979). We also performed a leave-one-out analysis; this procedure evaluates the robustness of the effect excluding one study at a time.

The meta-analyses and meta-regressions performed, along with the related plots, were computed using the R-package *metafor* (Viechtbauer, 2010).

Results

Moral decision-making domain

The random effect analysis ($N = 38$) showed a small to medium mean effect size, $g = .22$, 95% CI [.14, .30], $z = 5.44$, $p < .0001$, meaning that, compared with the NL, the FL induces participants to be more willing to accept harm in order to maximize outcomes. Total heterogeneity was significant, $Q_T = 109.19$, $p < .0001$, $I^2 = 64.74\%$, suggesting moderate variance across studies (see Fig. 1). However, it should also be noted that Higgins’s I^2 value can be considered medium. The Rosenthal’s fail-safe number was 1,175.

The leave-one-out analysis showed that the effect size was highly robust and ranged between .20 and .23 ($M = .22$, $SD = .006$). The trim-and-fill method added one hypothetical missing study on the left side of the funnel plot (see Fig. 2). Adding this hypothetical study, the effect size remained similar, $g = .21$, 95% CI [.13, .29], $z = 5.35$, $p < .0001$, and the heterogeneity remained significant, $Q_T = 111.20$, $p < .0001$, $I^2 = 64.30\%$. The Egger’s test was not significant, $z = .97$, $p = .32$, suggesting no publication bias.

The meta-regression performed using participants’ proficiency in the FL as moderator ($N = 38$) showed that the test on the moderator was not significant,³ $\chi^2(1) = .77$, $p = .37$, $\beta = -.004$; heterogeneity remained significant, $Q_T = 109.17$, $p < .0001$, $I^2 = 65.21\%$.

³ We also performed the same analysis using z scores. The effect of the moderator was still not significant, $\chi^2(1) = .77$, $p = .37$, $\beta = -.036$.

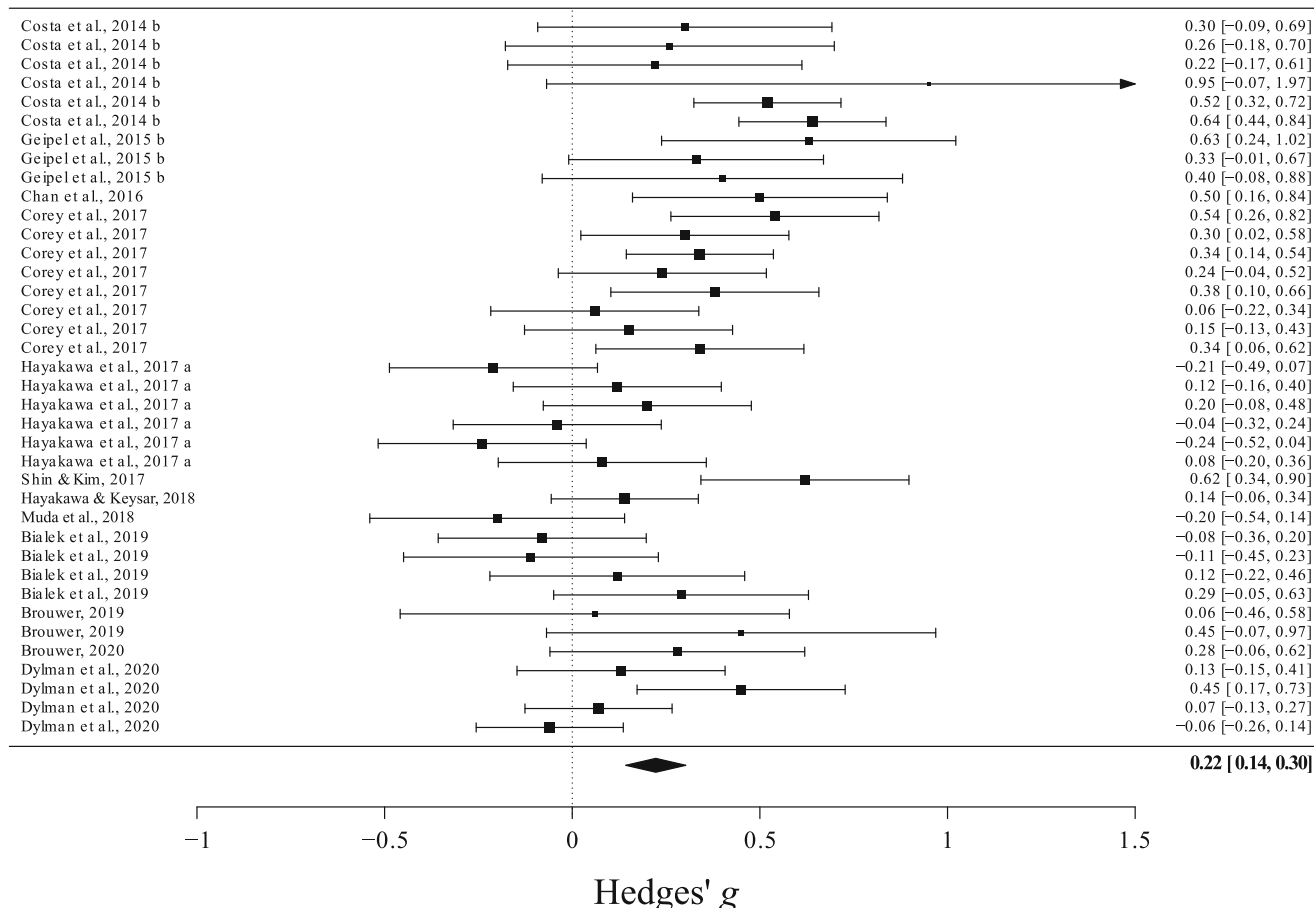


Fig. 1 Random effect analysis on studies assessing moral decision-making. For each study, Hedges's g and confidence intervals are reported. The cumulative results showed a small to medium effect size

The meta-regression performed using NL–FL similarity as moderator ($N = 38$) showed that the test on the moderator was significant, $\chi^2(1) = 10.35$, $p = .001$ and accounted for 37.09% of the heterogeneity which, however, remained significant, $Q_T = 76.82$, $p < .0001$, $I^2 = 53.35\%$. The decrease in heterogeneity suggests that NL–FL similarity plays a role in determining the differences in the effects reported by the various studies.

Critically, when the NL and FL belonged to different linguistic branches (i.e., Spanish vs. English), the effect size was significantly different from zero (see Table 2), $g = .30$, 95% CI [.21, .39], $z = 6.91$, $p < .0001$; on the other hand, when the NL and FL belonged to the same linguistic branch (i.e., German vs. English), the effect size was not significantly different from zero, $g = .06$, 95% CI [-0.05, .18], $z = -1.07$, $p = .28$.

Risk-aversion domain

The random effect analysis ($N = 9$) showed a small to medium mean effect size, $g = .28$, 95% CI [.17, .39], $z = 5.19$, $p < .0001$, meaning that an FL induces less risk aversion as compared with an NL. Total heterogeneity was not significant, $Q_T = 5.90$, $p = .82$, $I^2 = 0\%$,

suggesting that the studies reported similar results (see Fig. 3). The Rosenthal's fail-safe number was 77.

The leave-one-out analysis showed that the effect size was highly robust and ranged between .26 and .31 ($M = .28$, $SD = .01$). The trim-and-fill method added no hypothetical missing studies on the left side of the funnel plot (see Fig. 4). The Egger's test was not significant, $z = .37$, $p = .70$, suggesting no publication bias.

Discussion

The present study showed that an FL affects participants' decisions as compared with an NL in both the moral decision-making and risk-aversion domains: Being in an FL context made participants more willing to accept harm in order to maximize outcomes and reduced their risk aversion. The cumulative effects calculated for the moral decision-making domain were robust, but also moderately heterogeneous. Heterogeneity decreased after adding NL–FL similarity as a moderator in the meta-regression, suggesting that NL–FL similarity plays a role in determining the differences in the effects reported by the various studies. On the other hand, with regard

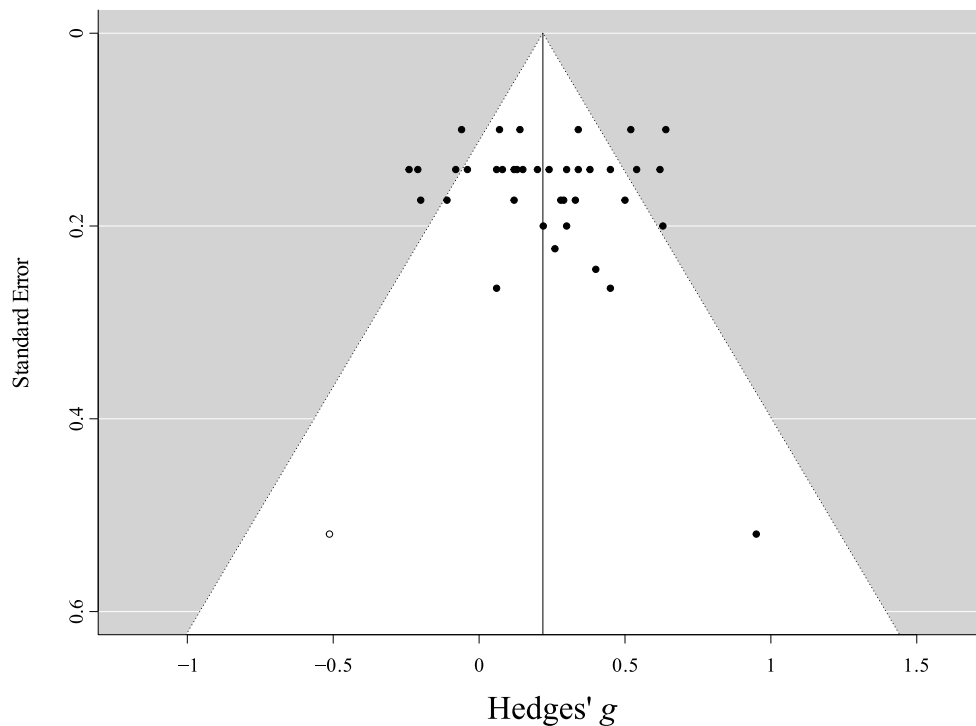


Fig. 2 Funnel plot computed using the trim-and-fill method on studies assessing moral decision-making. White dots indicate the hypothetical missing study. Adding this study, the effect size remained small to medium, revealing limited publication bias

to the risk-aversion domain, the effect calculated was robust and the heterogeneity was low, reflecting homogeneity in the task used and in the reported results.

The effect size of the FLE in the included studies varied between small and moderate, both in moral decision-making and in risk-aversion domains, confirming previous findings on the FLE. Critically, the present study also shows how the FLE literature is not distorted by a publication bias.

In the moral decision-making domain, results indicate a moderate heterogeneity, probably due to several factors such as participants' personality and individual differences, the NL and FL employed (i.e., degree of similarity/difference between them) and the difference in instructions and materials (i.e., in the risk-aversion domain the tasks involve mathematical reasoning to a large extent, while in the moral decision-making domain the task is more linguistically based). The two indices adopted to evaluate heterogeneity suggest different patterns, but the Q test is likely to be overpowered when the number of studies included is high. Thus, Higgins's I^2 should be a more reliable index of heterogeneity (Higgins et al.,

2003). Within this perspective, we note that in the meta-analyses reported herein Higgins's I^2 was always below the threshold of high heterogeneity (i.e., $I^2 > 75\%$; Higgins et al., 2003) and that adding NL–FL similarity as moderator accounted for a large part of this. Critically, when the NL and FL belonged to different linguistic branches (i.e., Spanish vs. English) the effect size was significantly different from zero, and when the NL and FL belonged to the same linguistic branch (i.e., German vs. English), the effect was null, consistent with recent evidence (Dylman & Champoux-Larsson, 2020; Miozzo et al., 2020).

In contrast, our findings also indicate that FL proficiency does not moderate the observed effect. This lack of effect could seem surprising, given the results available in the literature (for a review, see Pavlenko, 2017). However, it should be noted that herein we could only evaluate proficiency from a global level (i.e., the mean proficiency of each sample in each experiment), thus reducing the variability of hundreds of participants to one mean value. This approach is likely underpowered, and experimental studies employing mixed effects

Table 2 Hedges's g calculated using NL–FL similarity as moderator

	Different linguistic branches	Same linguistic branch
Moral decision-making	.30	.06
	[.21, .39]	[-.05, .18]
	$N = 25$	$N = 13$

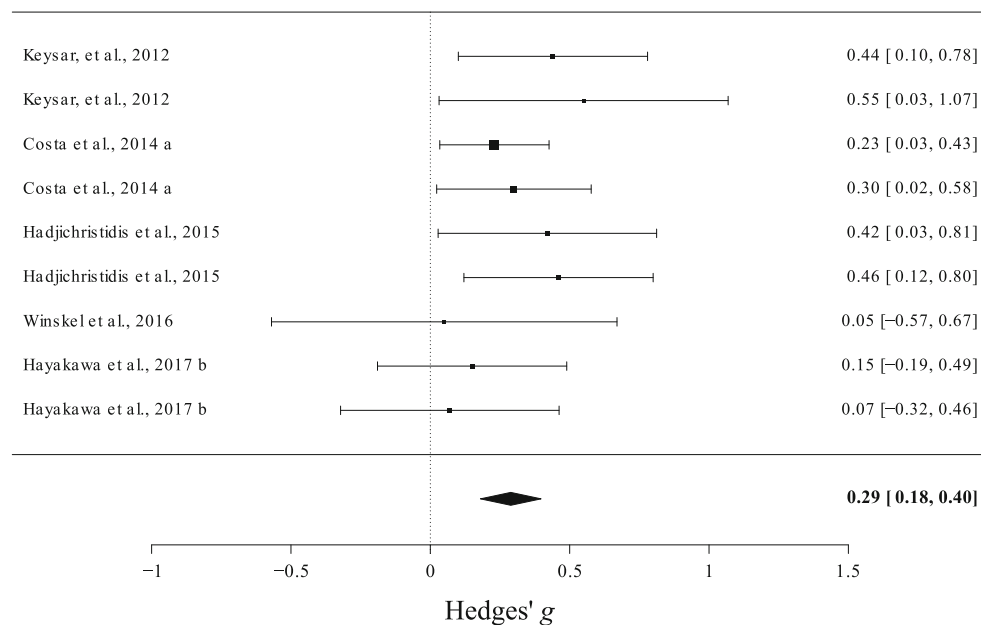


Fig. 3 Random effect analysis on studies assessing risk aversion. For each study, Hedges's' g and confidence intervals are reported. The cumulative results showed a small to medium effect size

linear/logistic regressions would better account for participants' variability in this factor.

In the domain of risk aversion, heterogeneity did not reach significance. This result is probably due to the greater similarity between the methodologies used in the studies that investigated the FLE in this domain. The effect size is very similar in the domain of moral decision-making and in that of risk

aversion. This may suggest similar or identical cognitive mechanisms underlying the FLE in those domains. Indeed, the literature suggests that emotions have a pervasive role in both domains (e.g., Hogarth, Portell, Cuxart, & Kolev, 2011; Youssef et al., 2012) and that emotions are more embodied in the NL (Pavlenko, 2012). However, such a conclusion could only be further justified through studies that administer tasks,

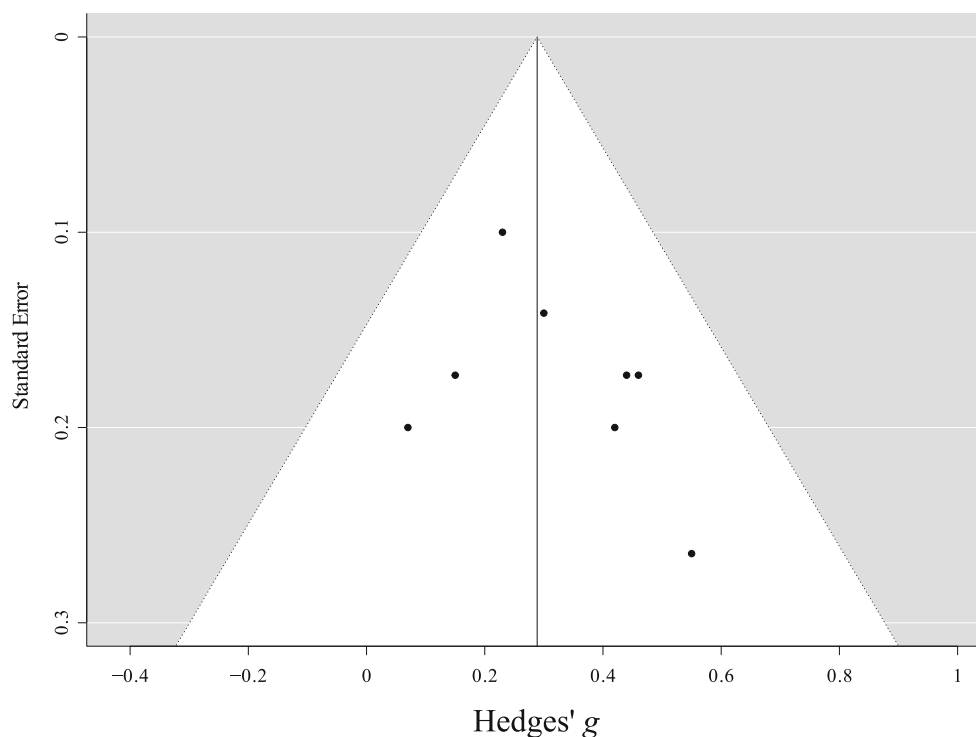


Fig. 4 Funnel plot computed using the trim-and-fill method on studies assessing risk aversion. The white dot indicates the hypothetical missing study. Adding this study, the effect size remained small to medium, revealing limited publication bias

both in the domain of moral decision-making and in the domain of risk aversion across the same participants, and then comparing the effects.

Taken together, our results provide suggestions for future research to investigate whether the FLE can prove generalizable outside the laboratory environment. In many published studies (for a review, see Hadjichristidis et al., 2019) the implications that the FLE could have on political decisions made in those contexts in which most of the people use a lingua franca (i.e., the United Nations or European Parliament) have often been discussed. To conclude that language could impact the way these decisions are made, we need to take into consideration the effect size, the role of the context and the moderating effect of NL–FL linguistic similarity. Therefore, future studies might directly investigate whether the FLE is modulated by public contexts and meetings involving a large number of people from different countries discussing one topic. As discussed recently by Hadjichristidis et al. (2019), the various languages differ from each other in terms of both cultural factors and grammatical features, which could be reflected in a different cognitive processing of information (see also Chen, 2013). International meetings are contexts in which situations similar to those investigated in the laboratory rarely occur. Rather, they are contexts in which the participants speak many different NLs and one FL (usually English). Future research could investigate the moderating role of NL–FL similarity on the FLE more profoundly. Our result on the moderating role of NL–FL similarity could also be due to the different number of studies in each category.

Finally, several limitations of this study should be mentioned. The studies included in this meta-analysis are those that met our inclusion criteria. We have adopted these conservative criteria in order to include in the meta-analysis only those studies which were comparable in terms of tasks and paradigms. We could not assume that the FLEs on different judgments and biases share the same underlying mechanisms: Even taking into consideration only studies that assessed the willingness to violate a social norm in order to maximize outcomes, among those concerning moral decisions in general, we found a high heterogeneity. Nevertheless, our choice had to be necessarily driven by theoretical considerations. Further meta-analyses, conducted by other authors and with different inclusion criteria, would be useful in order to validate our conclusions.

In conclusion, FL affects participants' decisions compared with NL, inducing participants to be more willing to accept harms in order to maximize outcomes in the moral decision-making domain and reducing risk aversion in risk-aversion domain. This result is critical within both experimental and social perspectives, and NL–FL similarity could be a critical moderator. Our results support previous findings on the FLE and provide suggestions for future research.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-020-01871-z>.

Author contributions R.C. and D.G.: Conceptualization, methodology, formal analysis, writing—original draft preparation. V.R. and T.V.: Writing—review and editing, supervision. All authors approved the final version of the manuscript for submission.

Funding This work was supported by funding from the Italian Ministry of University and Research (PRIN 2017 No. 201755TKFE) and from the Italian Ministry of Health (Ricerca Corrente 2020) to T.V.

Data availability Data and study materials (HTML and R-file of analysis code), that are necessary and sufficient to reproduce analyses and data presentations reported in the present study can be found at: https://github.com/danielegatti23/Foreing_Language_Effect_metanalysis

References

- Bereby-Meyer, Y., Hayakawa, S., Shalvi, S., Corey, J. D., Costa, A., & Keysar, B. (2018). Honesty speaks a second language. *Topics in Cognitive Science*, 12(2), 632–643. <https://doi.org/10.1111/tops.12360>
- Bialek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, 85(July). <https://doi.org/10.1016/j.jesp.2019.103855>
- Bialek, M., & Fugelsang, J. (2019). No evidence for decreased foreign language effect in highly proficient and acculturated bilinguals: A commentary on Čavar and Tytus (2018). *Journal of Multilingual and Multicultural Development*, 40(8), 679–686. <https://doi.org/10.1080/01434632.2018.1547072>
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Wiley.
- Boucher, J., & Osgood C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 8, 1–8. [https://doi.org/10.1016/S0022-5371\(69\)80002-2](https://doi.org/10.1016/S0022-5371(69)80002-2)
- Brouwer, S. (2019). The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *Journal of Multilingual and Multicultural Development*, 40(10), 865–878. <https://doi.org/10.1080/01434632.2019.1585863>
- Brouwer, S. (2020). The interplay between emotion and modality in the foreign-language effect on moral decision making. *Bilingualism: Language and Cognition*, 1–8. Advance online publication. <https://doi.org/10.1017/S136672892000022X>
- Caldwell-Harris, C. L. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science*, 24(3), 214–219. <https://doi.org/10.1177/0963721414566268>
- Čavar, F., & Tytus, A. E. (2017). Moral judgement and foreign language effect: When the foreign language becomes the second language. *Journal of Multilingual and Multicultural Development*, 39(1), 17–28. <https://doi.org/10.1080/01434632.2017.1304397>
- Chan, Y. L., Gu, X., Ng, J. C. K., & Tse, C. S. (2016). Effects of dilemma type, language, and emotion arousal on utilitarian vs deontological choice to moral dilemmas in Chinese–English bilinguals. *Asian Journal of Social Psychology*, 19(1), 55–65. <https://doi.org/10.1111/ajsp.12123>
- Chen, M. K., (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2), 690–731. <https://doi.org/10.1257/aer.103.2.690>

- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, 29(1), 23–40. <https://doi.org/10.1080/09515089.2014.993063>
- 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. <https://doi.org/10.1037/a0031021>
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. <https://www.jstor.org/stable/10.7758/9781610448864>
- Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000356>
- Costa, A., Corey, J., Hayakawa, S., Aparici, M., Vives, M., & Keysar, B. (2018). The role of intentions and outcomes in the foreign language effect on moral judgements. *Quarterly Journal of Experimental Psychology*, 72(1), 1–10. <https://doi.org/10.1177/1747021817738409>
- Costa, A., Foucart, A., Armon, I., Aparici, M., & Apesteguia, J. (2014a). “Piensa” twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254. <https://doi.org/10.1016/j.cognition.2013.11.010>
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014b). Your morals depend on language. *PLOS ONE*, 9(4). <https://doi.org/10.1371/journal.pone.0094842>
- Del Re, A. C. (2013). compute.es: Compute effect sizes (R Package Version 0.2-2) [Computer software]. <http://cran.r-project.org/web/packages/compute.es>
- Del Re, A. C., & Hoyt, W. T. (2014). MAD: Meta-analysis with mean differences (R Package Version 0.8-2) [Computer software]. <https://cran.r-project.org/package=MAD>
- Dewaele, J. M. (2004). Perceived language dominance and language preference for emotional speech: The implications for attrition research. In M. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 81–104). John Benjamins. <https://doi.org/10.1075/sibil.28.06dew>
- Díaz-Lago, M., & Matute, H. (2018). Thinking in a Foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, 72(1), 41–51. <https://doi.org/10.1177/1747021818755326>
- Driver, M. (2020). Switching codes and shifting morals: How code-switching and emotion affect moral judgment. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2020.1730763>
- Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). Wiley.
- Dylman, A., & Champoux-Larsson, M. F. (2020). It’s (not) all Greek to me: Boundaries of the foreign language effect. *Cognition*, 196. <https://doi.org/10.1016/j.cognition.2019.104148>
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Foroni, F. (2015). Do we embody second language? Evidence for ‘partial’ simulation during processing of a second language. *Brain and Cognition*, 99, 8–16. <https://doi.org/10.1016/j.bandc.2015.06.006>
- Gao, S., Zika, O., Rogers, R. D., & Thierry, G. (2015). Second language feedback abolishes the “hot hand” effect during even-probability gambling. *Journal of Neuroscience*, 35(15), 5983–5989. <https://doi.org/10.1523/JNEUROSCI.3622-14.2015>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, 59, 8–17. <https://doi.org/10.1016/j.jesp.2015.02.001>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). The foreign language effect on moral judgment: The role of emotions and norms. *PLOS ONE*, 10(7), 1–17. <https://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. <https://doi.org/10.1016/j.cognition.2016.05.010>
- Geipel, J., Hadjichristidis, C., & Klesse, A. K. (2018). Barriers to sustainable consumption attenuated by foreign language use. *Nature Sustainability*, 1, 31–33. <https://doi.org/10.1038/s41893-017-0005-9>
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107, 1144–1154. <https://doi.org/10.1016/j.cognition.2007.11.004>
- Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied*, 21(2), 117–129. <https://doi.org/10.1037/xap0000044>
- Hadjichristidis, C., Geipel, J., & Keysar, B. (2019). The influence of native language in shaping judgment and choice. In N. Srinivasan (Ed.), *Progress in brain research* (Vol. 247, 253–272). Elsevier. <https://doi.org/10.1016/bs.pbr.2019.02.003>
- Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition*, 173(December 2017), 8–15. <https://doi.org/10.1016/j.cognition.2017.12.010>
- Hayakawa, S., Lau, B. K. Y., Holtzmann, S., Costa, A., & Keysar, B. (2017a). On the reliability of the foreign language effect on risk-taking. *Quarterly Journal of Experimental Psychology*, 72(1), 29–40. <https://doi.org/10.1177/1747021817742242>
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017b). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, 28(10), 1387–1397. <https://doi.org/10.1177/0956797617720944>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hogarth, R. M., Portell, M., Cuxart, A., & Kolev, I. M. (2011). Emotion and reason in everyday risk perception. *Journal of Behavioral Decision Making*, 24, 202–222. <https://doi.org/10.1002/bdm.689>
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541. [https://doi.org/10.1016/0749-596X\(91\)90025-F](https://doi.org/10.1016/0749-596X(91)90025-F)
- Jóńczyk, R., Korolczuk, I., Balatsou, E., Thierry, G. (2019). Keep calm and carry on: Electrophysiological evaluation of emotional anticipation in the second language. *Social Cognitive and Affective Neuroscience*, 14(8) 885–898. <https://doi.org/10.1093/scan/nsz066>
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668. <https://doi.org/10.1177/0956797611432178>
- Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, 74(3), 785–794. <https://doi.org/10.1111/biom.12817>
- Mækela, J. M., & Pfuhl, G. (2019). Deliberate reasoning is not affected by language. *PLOS ONE*, 14(1), e0211428. <https://doi.org/10.1371/journal.pone.0211428>
- Miozzo, M., Navarrete, E., Ongis, M., Mello, E., Giroto, V., & Peressotti, F. (2020). Foreign language effect in decision-making: How foreign is it?. *Cognition*, 199. <https://doi.org/10.1016/j.cognition.2020.104245>

- Muda, R., Niszczoła, P., Bialek, M., & Conway, P. (2018). Reading dilemmas in a foreign language reduces both deontological and utilitarian response tendencies. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(2), 321–326. <https://doi.org/10.1037/xlm0000447>
- Nadarevic, L., Plier, S., Thielmann, I., & Darancó, S. (2018). Foreign language reduces the longevity of the repetition-based truth effect. *Acta Psychologica*, 191(February), 149–159. <https://doi.org/10.1016/j.actpsy.2018.08.019>
- Oganian, Y., Korn, C. W., & Heekeren, H. R. (2016). Language switching—but not foreign language use per se—reduces the framing effect. *Journal of Experimental Psychology: Learning Memory and Cognition*, 42(1), 140–148. <https://doi.org/10.1037/xlm0000161>
- 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. <https://doi.org/10.1080/01434630408666528>
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6), 405–428. <https://doi.org/10.1080/00207594.2012.743665>
- Pavlenko, A. (2017). Do you wish to waive your rights?: Affect and decision-making in multilingual speakers. *Current Opinion in Psychology*, 17. <https://doi.org/10.1016/j.copsyc.2017.06.005>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638. <https://doi.org/10.1037/0033-2909.86.3.638>
- RStudio Team. (2015). RStudio: Integrated development for R [Computer software]. <http://www.rstudio.com/>. Accessed: August 2020
- Sheikh, N. A., & Titone, D. (2016). The embodiment of emotional words in a second language: Aneye-movement study. *Cognition & Emotion*, 30(3), 488–500. <https://doi.org/10.1080/02699931.2015.1018144>
- Shin, H. I., & Kim, J. (2017). Foreign language effect and psychological distance. *Journal of Psycholinguistic Research*, 46(6), 1339–1352. <https://doi.org/10.1007/s10936-017-9498-7>
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–665. <https://doi.org/10.1017/CBO9780511808098.026>
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458. <https://doi.org/10.1126/science.7455683>
- Van Hugten, J., & van Witteloostuijn, A. (2018). The foreign language effect on the self-serving bias: A field experiment in the high school classroom. *PLOS ONE*, 13(2), 1–16. <https://doi.org/10.1371/journal.pone.0192143>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <http://www.jstatsoft.org/v36/i03/>. Accessed: August 2020
- Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behaviour*. Princeton University Press.
- Winkel, H., Ratitamkul, T., Brambley, V., Nagarachinda, T., & Tiencharoen, S. (2016). Decision-making and the framing effect in a foreign and native language. *Journal of Cognitive Psychology*, 28(4), 427–436. <https://doi.org/10.1080/20445911.2016.1139583>
- Wong, G., & Ng, B. C. (2018). Moral judgement in early bilinguals: Language dominance influences responses to moral dilemmas. *Frontiers in Psychology*, 9, 1–10. <https://doi.org/10.3389/fpsyg.2018.01070>
- Wu, Y. J., & Thierry, G. (2012). How reading in a second language protects your heart. *The Journal of Neuroscience*, 32(19), 6485–6489. <https://doi.org/10.1523/JNEUROSCI.6119-11.2012>
- Wu, Y. J., & Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *The Journal of Neuroscience*, 33(33), 13533–13537. <https://doi.org/10.1523/JNEUROSCI.4760-12.2013>
- Youssef, F. F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., Maloo, S., Degannes, J., Dobo, L., Ditshotlo, P., & Legall, G. (2012). Stress alters personal moral decision making. *Psychoneuroendocrinology*, 37(4), 491–498. <https://doi.org/10.1016/j.psyneuen.2011.07.017>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.