Thinking More or Feeling Less? Explaining the Foreign-Language Effect on Moral Judgment

Sayuri Hayakawa[1], David Tannenbaum[2], Albert Costa[3,4], Joanna D. Corey[3], and Boaz Keysar[1]
[1]Department of Psychology, University of Chicago; [2]David Eccles School of Business, University of Utah; [3]Center of Brain and Cognition (CBC), Universitat Pompeu Fabra; and [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract

Would you kill one person to save five? People are more willing to accept such utilitarian action when using a foreign language than when using their native language. In six experiments, we investigated why foreign-language use affects moral choice in this way. On the one hand, the difficulty of using a foreign language might slow people down and increase deliberation, amplifying utilitarian considerations of maximizing welfare. On the other hand, use of a foreign language might stunt emotional processing, attenuating considerations of deontological rules, such as the prohibition against killing. Using a process-dissociation technique, we found that foreign-language use decreases deontological responding but does not increase utilitarian responding. This suggests that using a foreign language affects moral choice not through increased deliberation but by blunting emotional reactions associated with the violation of deontological rules.

Would you kill one person to save five? Decisions such as these can be difficult because they pit common moral rules ("do not actively harm innocent persons") against a desire to maximize welfare (in this case, by saving as many lives as possible). In this way, moral dilemmas embody a tension between deontological prescriptions that forbid certain behaviors regardless of the consequences (Kant, 1785/1959) and utilitarian prescriptions concerned with bringing about the greatest good for the greatest number of people (Mill, 1861/1998).

Responses to moral dilemmas often depend on a number of contextual factors, such as the decision maker's relationship to the involved parties or how the action would be carried out. Remarkably, recent discoveries have shown that responses to moral dilemmas also systematically depend on whether these decisions are made in a foreign or native language. Several studies have found that bilingual speakers are more likely to endorse what appear to be utilitarian moral decisions[1]

when responding in a foreign language than when responding in their native language (Cipolletti, McFarlane, & Weissglass, 2016; Corey et al., 2017; Costa et al., 2014; Geipel, Hadjichristidis, & Surian, 2015a). In one study (Costa et al., 2014), bilinguals considered the classic footbridge dilemma, in which five people tied to a train track are about to be killed by an oncoming trolley (Foot, 1978; Thomson, 1985). The only way to save them would be to push a large bystander onto the tracks, thereby killing him but stopping the train. Only 18% of participants were willing to sacrifice the large man when the problem was presented in their native language, whereas 44% were willing to do so when it was presented in their foreign language. This *moral*

Corresponding Author:
Sayuri Hayakawa, University of Chicago, Department of Psychology, 5848 S. University Ave., Chicago, IL 60637
E-mail: sayuri@uchicago.edu

foreign-language effect (MFLE) was found with native English, Hebrew, and Korean speakers who spoke French, Spanish, or English as a foreign language. Other research teams have independently replicated these results with different bilingual populations (Cipolletti et al., 2016; Geipel et al., 2015a).

Although the MFLE appears to be robust across a variety of languages, it is unclear why foreign-language use affects moral judgment. Here, we adopt a dual-process framework (Stanovich & West, 2000) as a tool to consider possible mechanisms underlying this phenomenon. According to this framework, decisions are made through the interplay of at least two systems, one involving mental processes that are relatively quick, effortless, and intuitive (System 1) and another involving mental processes that are relatively slow, effortful, and deliberative (System 2; Epstein, 1994; Kahneman, 2003). Although any given decision may be the result of either of these systems, there is evidence that behaviors consistent with deontological, rule-based proscriptions, such as "do not actively cause harm," are preferentially supported by System 1 processes, whereas utilitarian judgments, such as "maximize the greatest good for the greatest number of people," are supported by System 2 (e.g., Cushman, 2013; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008). Responding to moral dilemmas in a foreign language could affect choice by perturbing either one or both of these systems.

We compare two theoretical explanations of the MFLE. One possibility, the *blunted-deontology* account, holds that foreign-language use affects moral decisions by stunting the emotional or heuristic processing characteristic of System 1. For instance, when people read taboo words in a foreign language rather than their native language, they rate those words as less emotionally evocative and their physiological responses are weaker (e.g., Harris, Ayçiçeği, & Gleason, 2003; Puntoni, de Langhe, & van Osselaer, 2009). Additionally, consumers tend to make more emotional or hedonic choices when indicating their preferences through speech rather than other modalities, such as pointing, and such differences disappear when they use a foreign language (Klesse, Levav, & Goukens, 2015). Thus, a reduction in emotional processing when participants use a foreign language may increase their willingness to sacrifice one person to save five because doing so is not especially aversive, or because moral rules are not particularly salient (see Geipel, Hadjichristidis, & Surian, 2015b).

A second possibility, the *heightened-utilitarianism* account, posits that foreign-language use affects moral decisions by encouraging deliberative thinking characteristic of System 2. Responding in a foreign language is often cognitively effortful and can increase feelings of processing difficulty (i.e., metacognitive disfluency),

which in turn promote greater analytic thinking (Oppenheimer, 2008; Schwarz, 2011). Thus, the disfluency participants experience when using a foreign language could prompt them to engage in more deliberative System 2 thinking associated with utilitarian judgment. According to this account, foreign-language users, compared with native-language users, do not necessarily feel less averse to sacrificing an innocent person to save five, but instead are likely to place greater weight on maximizing net welfare.

Just as dual-process models posit that System 1 and System 2 processes are psychologically independent of one another (e.g., Wang, Highhouse, Lake, Petersen, & Rada, 2017), the blunted-deontology and heightened-utilitarianism accounts are conceptually independent explanations of the MFLE. However, as first noted by Conway and Gawronski (2013), methods commonly used by moral psychologists do not separate deontological from utilitarian responses. For example, a participant's willingness to sacrifice one life in order to save five is typically taken as an indication of both increased utilitarian and decreased deontological responding. Because utilitarian responding and deontological responding have been treated as two ends of a single continuum, existing research cannot shed light on whether foreign-language use affects moral decisions by heightening utilitarian considerations, blunting deontological considerations, or both.

This conceptual confusion has led independent research teams to interpret the same pattern of results differently. For instance, Costa et al. (2014, abstract) suggested that the MFLE is due to a reduction in the "emotional response elicited by the foreign language," which reduces "the impact of intuitive emotional concerns." However, Cipolletti et al. (2016) explained the MFLE by positing that "thinking in one's non-native language activates systematic processing characteristic of [System 2] processing" (p. 26). The fact that researchers have drawn different conclusions from similar findings highlights the ambiguity of the behavioral data, and the need to identify the mechanism underlying the MFLE.

The goal of our research was to experimentally separate deontological responding indicative of System 1 and utilitarian responding characteristic of System 2 in order to understand how foreign-language use affects moral judgment. We conducted six experiments utilizing a process-dissociation technique that disentangles utilitarian and deontological judgment (Conway & Gawronski, 2013; Jacoby, 1991).

General Method

Our experiments varied in the language populations studied and the experimental stimuli used, but all

shared the same basic procedure. For purposes of efficiency, we first describe the basic experimental procedure and then discuss differences among the experiments. Sample sizes were always determined in advance, and we report all data exclusions and manipulations. These six experiments represent every study we have conducted to test our hypothesis (i.e., our entire "file drawer").

Participants

For each experiment, we targeted a final sample size of 200 participants (100 per language condition) to obtain groups comparable in size to those utilized by Conway and Gawronski (2013). Sample sizes varied somewhat because of our exclusion criteria, which we discuss in the Procedure section. Participants' payment ranged from €2 to €5 or $2 to $5, depending on the experiment. All participants were bilingual, and most had acquired their foreign language in a classroom setting. None of our participants grew up speaking their foreign language at home. Table 1 provides sample sizes for all the experiments.

Procedure

Participants were randomly assigned to complete the study in either their native or their foreign language. All experiments except Experiment 3 were conducted online; Experiment 3 was conducted in a laboratory setting in Barcelona, Spain, by a bilingual experimenter.

In all cases, the experiment was administered entirely in the assigned language. All materials were translated and back-translated from English to ensure comparability (Brislin, 1970).

We followed the same screening and exclusion criteria that we have used in our past work on foreign-language effects (Costa et al., 2014; Keysar, Hayakawa, & An, 2012). After completing an initial language-background screening, participants were allowed to proceed to the study only if they reported (a) being a native speaker of the target native language, (b) being a foreign speaker of the target foreign language, and (c) not growing up speaking the target foreign language at home. Eligible participants went through a second phase of screening by completing a proficiency quiz that involved reading a paragraph in the assigned language and answering a multiple-choice question about what they had just read. Only participants who answered the question correctly were allowed to participate in the experiment.

After the screening procedure, participants were presented with 20 moral dilemmas (see the next section for details). After completing this process-dissociation task, participants in Experiments 1 and 2 also completed three short individual difference measures, which we included for exploratory purposes and also to serve as a replication of Conway and Gawronski (2013). These measures were a seven-item subscale of the Interpersonal Reactivity Index that measures general empathic concern towards other individuals (Davis, 1983), a five-item scale measuring general need for cognition (Need

Table 1. Overview of the Experiments

Experiment	Sample size	Native language	Foreign language	Moral dilemmas	Elicitation format
Experiment 1	214	German	English	Original stimuli from Conway and Gawronski (2013)	Judgment (e.g., "Is it appropriate to push the man off the bridge?")
Experiment 2	242	English	Spanish	Original stimuli from Conway and Gawronski (2013)	Judgment (e.g., "Is it appropriate to push the man off the bridge?")
Experiment 3	195	Spanish	English	Revised stimuli from Conway and Rosas (2017)	Judgment with consequences highlighted (e.g., "Is it morally correct to push the man off the bridge to save five people, even though the man would die?")
Experiment 4	211	German	English	Revised stimuli from Conway and Rosas (2017)	Judgment with consequences highlighted (e.g., "Is it morally correct to push the man off the bridge to save five people, even though the man would die?")
Experiment 5	209	German	English	Revised stimuli from Conway and Rosas (2017)	Choice (e.g., "Would you push the man off the bridge to save five people?")
Experiment 6	206	English	German	Revised stimuli from Conway and Rosas (2017)	Choice (e.g., "Would you push the man off the bridge to save five people?")

for Cognition Scale; Cacioppo, & Petty, 1982), and a five-item measure of cognitive reflection (Cognitive Reflection Test; Baron, Scott, Fincher, & Metz, 2014; Frederick, 2005). Results from these individual difference measures are largely consistent with those reported by Conway and Gawronski (2013) and are reported in the Supplemental Material available online.

Next, participants translated one moral dilemma from the designated language of the experiment to the other language. This was done to ensure that participants had attended to and comprehended the target stimulus materials. The dilemma was randomly selected in advance from the stimulus set and was the same for all participants in a given experiment. Non-English translations were translated back into English using Google Translate, checked by a native English speaker, and then confirmed by a native speaker of the original language. English translations were checked by a native English speaker. Participants who failed to translate any part of the dilemma or who wrote gibberish were excluded from the final analysis.

Finally, participants completed a set of demographic questions and rated their proficiency speaking, listening, reading, and writing in their native and foreign languages. Table S1 in the Supplemental Material presents summary statistics for these measures.

Process-dissociation task

The order of the 20 moral dilemmas was randomized for each participant. For each scenario, participants either determined whether a given action was appropriate (Experiments 1–4) or determined if they would be willing to perform the action themselves (Experiments 5 and 6). The response options were "yes," "no," and "I don't understand." We excluded trials in which participants selected the "I don't understand" option (0.26–1.54% of trials across studies).

In each experiment, we used a set of moral dilemmas designed to provide independent measures of deontological and utilitarian responding for each participant (Conway & Gawronski, 2013). The key feature of this technique is the presentation of 10 incongruent and 10 congruent moral dilemmas. Traditional moral dilemmas, such as sacrificing one person to save five people, are incongruent in the sense that deontological and utilitarian concerns conflict: Deontological concerns prohibit killing a person to save five, whereas utilitarian concerns demand it. Congruent dilemmas are structurally identical to incongruent dilemmas except that deontological and utilitarian considerations are in agreement. For example, if the choice concerns sacrificing one life in order to prevent five people from being mildly

injured, neither deontological nor utilitarian concerns would endorse sacrificing the one person. Each participant responded to 10 pairs of congruent and incongruent dilemmas.

Comparing response rates for congruent and incongruent dilemmas allowed us to recover separate measures of deontological and utilitarian responding. To do this, we followed the method detailed in Conway and Gawronski (2013). First, we calculated a utilitarianism parameter (U) for each participant by taking the difference in the proportion of "no" responses between congruent trials and incongruent trials:

$$U = p(\text{unacceptable} \mid \text{congruent}) - p(\text{unacceptable} \mid \text{incongruent}). \quad (1)$$

Thus, participants scoring high on utilitarianism found harmful actions unacceptable when they failed to maximize net welfare (i.e., congruent trials), but acceptable when they maximized net welfare (i.e., incongruent trials). Those scoring near 0 on this measure judged harmful actions as comparably acceptable regardless of whether the actions maximized net welfare. Scores could range from −1 to 1, but the mass of the distribution fell between 0 and 1 (negative U scores were possible but rare, as they would imply that participants found it acceptable to kill an innocent person to save five from mild harm, but not acceptable to kill an innocent person to save five lives).[2]

To arrive at a separate measure of deontological considerations (D), we determined the proportion of instances in which utilitarianism did not drive responses ($1 - U$). This term includes judgments driven by deontological considerations plus any other response tendency to find actions acceptable in both congruent and incongruent trials. To isolate D, we calculated the proportion of "no" responses in incongruent trials relative to all nonutilitarian responses:

$$D = p(\text{unacceptable} \mid \text{incongruent})/(1 - U). \quad (2)$$

Thus, D can be thought of as what is left over after partialing out both nonutilitarian and nondeontological response tendencies. Scores on D could range from 0 to 1; higher scores indicated greater deontological responding.[3]

Experimental Permutations

Our six experiments differed along three dimensions: (a) the type of bilingual population used, (b) how we elicited responses from participants, and (c) the set of

moral dilemmas participants responded to. We discuss each permutation in this section; Table 1 provides an overview.

Bilingual populations

Both participants' native language and their foreign language varied across the six experiments. Participants' native language was either German (Experiments 1, 4, and 5), English (Experiments 2 and 6), or Spanish (Experiment 3). Participants' foreign language was either German (Experiment 6), English (Experiments 1, 3, 4, and 5), or Spanish (Experiment 2). Collectively, our experiments allowed us to examine whether effect sizes varied according to specific native or foreign languages. In some cases, we could also compare experiments in which native and foreign languages were fully crossed. For example, we could compare the MFLE of native German speakers responding in English (Experiment 5) with the MFLE of native English speakers responding in German (Experiment 6). Doing so allowed us to cleanly disentangle whether our results were driven by using a foreign language in general rather than by using a specific foreign language.

Elicitation format

Our experiments differed in how participants provided their responses. In Experiments 1 and 2, participants were asked questions of the form "Is it appropriate to push the man off the bridge?" This is similar to the elicitation format used by Conway and Gawronski (2013). For Experiments 3 and 4, participants were asked questions of the form "Is it morally correct to push the man off the bridge to save five people, *even though the man would die*?" (emphasis added here). This allowed us to examine whether highlighting the negative consequences of engaging in utilitarian action would affect our results. In Experiments 5 and 6, participants were asked questions of the form "Would you push the man off the bridge to save five people?" This elicitation format allowed us to examine whether using a foreign language affects both moral judgment and choice.

Moral dilemmas

Our experiments used two different sets of moral dilemmas. Experiments 1 and 2 used the original set of dilemmas from Conway and Gawronski (2013), whereas Experiments 3 through 6 used an updated set of dilemmas (Conway & Rosas, 2017). Both sets comprised 20 scenarios and were designed to recover separate U and D parameters for each participant, but the

sets differed in their content. This allowed us to examine the robustness of the MFLE across a range of different situations.

Results

We recovered a single U parameter and a single D parameter for each participant from his or her choices, according to Equations 1 and 2. Conway and Gawronski (2013) did not observe a reliable correlation between U and D scores, and we replicated that finding in all of our experiments with the exception of Experiment 3 ($r = -.20$, $p = .007$; see Table S3 in the Supplemental Material for more details). When we restricted our analysis to the dilemmas used in the incongruent trials (which were similar to the dilemmas traditionally used to measure utilitarianism), responses in all six experiments were positively correlated with U and negatively correlated with D (all ps < .001). These results provide empirical support that both U and D scores predict traditional measures of utilitarianism but that the two parameters are statistically independent of one another.

Primary analysis

We report the results from a meta-analysis of the data across our six experiments. For this analysis, we modeled experiment as a random effect (Lipsey & Wilson, 2001). Treating experiment instead as a fixed effect returned even stronger results.

Overall, our results are consistent with the blunted-deontology account (see Table 2). We consistently observed lower D scores for participants in the foreign-language condition compared with participants in the native-language condition; the overall effect size, d, was 0.24 ($p < .001$).[4] Figure 1 shows the standardized mean difference between conditions for each of the six experiments, along with the aggregate mean-difference score from all six experiments. Although the magnitude of the language effect varied somewhat from study to study, in all six experiments, participants using their native language responded more deontologically than did those using a foreign language.

In contrast, we failed to find support for the heightened-utilitarianism account (see Fig. 1 and Table 2). In no experiment did we find a reliable increase in utilitarianism among participants responding in their foreign language, and in three experiments, we observed a reliable *decrease* in utilitarianism for participants in the foreign-language condition. Across our experiments, participants in the foreign-language condition had lower U scores compared with participants in the native-language condition (combined $d = 0.25$, $p = .022$). These

Table 2. Comparison of Responding in the Foreign-Language and Native-Language Conditions

Experiment	Deontological responding (D)				Utilitarian responding (U)				Traditional utilitarianism ("yes" responses to incongruent dilemmas)			
	Native-language condition	Foreign-language condition	Cohen's d	p	Native-language condition	Foreign-language condition	Cohen's d	p	Native-language condition	Foreign-language condition	Cohen's d	p
Experiment 1	0.72 (0.18)	0.65 (0.21)	0.362	.009	0.34 (0.19)	0.33 (0.20)	0.063	> .250	0.52 (0.19)	0.56 (0.19)	−0.201	.144
Experiment 2	0.67 (0.20)	0.61 (0.19)	0.304	.019	0.36 (0.18)	0.26 (0.18)	0.578	< .001	0.57 (0.17)	0.55 (0.17)	0.131	.310
Experiment 3	0.76 (0.29)	0.75 (0.25)	0.042	> .250	0.37 (0.25)	0.28 (0.24)	0.381	.009	0.50 (0.25)	0.45 (0.25)	0.217	.169
Experiment 4	0.83 (0.25)	0.80 (0.24)	0.130	> .250	0.23 (0.24)	0.24 (0.27)	−0.025	> .250	0.36 (0.28)	0.37 (0.26)	−0.047	> .250
Experiment 5	0.78 (0.26)	0.70 (0.30)	0.290	.038	0.32 (0.23)	0.32 (0.27)	0.015	> .250	0.47 (0.24)	0.53 (0.26)	−0.223	.107
Experiment 6	0.78 (0.29)	0.70 (0.26)	0.300	.033	0.43 (0.26)	0.31 (0.24)	0.501	< .001	0.54 (0.27)	0.52 (0.23)	0.061	> .250
Combined results			0.240	< .001			0.252	.022			−0.010	> .250

Note: Values in parentheses are standard deviations. The p values are from t tests comparing responding in the two language conditions. Positive difference (Cohen's d) scores for D would be consistent with the blunted-deontology account, and negative difference scores for U would be consistent with heightened utilitarianism.
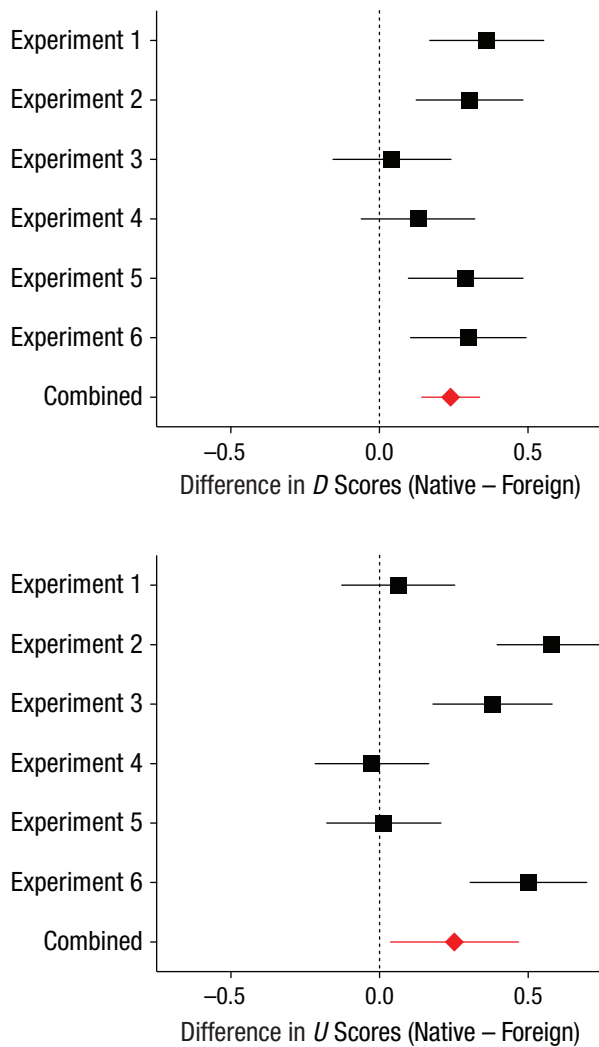
Fig. 1. Forest plots of the results from Experiments 1 through 6. The graphs plot the standardized mean difference (i.e., Cohen's *d*) in *D* scores (top panel) and *U* scores (bottom panel) between the native-language and foreign-language conditions, along with combined effect sizes across all the experiments, calculated using a random-effects model (Lipsey & Wilson, 2001). Positive numbers indicate higher scores in the native-language condition relative to the foreign-language condition. Positive values for the difference in *D* scores would support the blunted-deontology account, and negative values for the difference in *U* scores would support the heightened-utilitarianism account. Error bars represent 95% confidence intervals.

findings are in direct opposition to the idea that foreign-language use increases utilitarianism.

We also examined how foreign-language use affected responses specifically to the incongruent dilemmas. As Table 2 shows, we did not observe significant results in any of the experiments individually or when the data were aggregated across experiments, combined *d* = −0.010 (*p* > .250). This null effect is inconsistent with previous findings (e.g., Corey et al., 2017; Costa et al., 2014; Geipel et al., 2015a), but we note that our

dilemmas differed from those used in earlier studies. In addition, unlike previous studies, ours directly juxtaposed incongruent and congruent dilemmas, which could have led to contrast effects.

Planned contrast tests

We next conducted a series of planned orthogonal contrasts (Furr & Rosenthal, 2003) to more directly test and discriminate between the blunted-deontology account and the heightened-utilitarianism account. Using planned contrasts also allowed us to test a hybrid account according to which foreign-language use both blunts deontological reasoning and heightens utilitarian reasoning. First, we standardized *U* and *D* scores to have a mean of 0 and standard deviation of 1 in order to remove arbitrary differences in how the two parameters are scaled. According to the blunted-deontology account, *D* scores should be lower when participants respond in a foreign language (L2) than when they respond in their native language (L1), and there should be no differences in *U* scores between the two conditions: $D_{L2} < D_{L1} \approx U_{L1} \approx U_{L2}$. According to the heightened-utilitarianism account, *U* scores should be higher when participants respond in a foreign language then when they respond in their native language, and there should be no differences in *D* scores between the two conditions: $D_{L2} \approx D_{L1} \approx U_{L1} < U_{L2}$. According to the hybrid account, both scores should differ between the two conditions: $D_{L2} < D_{L1} \approx U_{L1} < U_{L2}$. We constructed orthogonal contrasts representing these predictions (Rosenthal & Rosnow, 1985) and regressed participants' *U* and *D* scores (calculated using Equations 1 and 2) onto each contrast separately.[5] Contrasts were coded such that positive coefficients indicated support for a given hypothesis.

As in our earlier analysis, we found clear support for the blunted-deontology account (see Table 3). In all six experiments, we observed a positive coefficient (i.e., an effect in the predicted direction) for the blunted-deontology contrast, and the fit to the data was reliable when we combined results across experiments (again using a random-effects model), combined *b* = 0.040, 95% confidence interval (CI) = [0.018, 0.062], *p* < .001. On the other hand, in no experiment did we find a reliable effect in the predicted direction for the heightened-utilitarianism contrast, and in some cases, the contrasts were statistically significant in the opposite direction, mirroring the results of our primary analysis (see Table 3). Finally, our contrast testing the hybrid account did not yield significant results when the experiments were tested individually or when the data were combined across experiments, combined *b* = −0.010, 95% CI = [−0.065, 0.044], *p* > .250 (Table 3).

Table 3. Results From the Orthogonal Contrast Tests

Experiment	Contrast 1: blunted-deontology account			Contrast 2: heightened-utilitarianism account			Contrast 3: hybrid account		
	b	95% CI	p	b	95% CI	p	b	95% CI	p
Experiment 1	**0.056**	[0.001, 0.111]	.047	−0.010	[−0.065, 0.045]	> .250	**0.068**	[−0.069, 0.205]	> .250
Experiment 2	**0.050**	[0.001, 0.100]	.047	−0.093	[−0.142, −0.044]	< .001	−0.064	[−0.181, 0.053]	> .250
Experiment 3	**0.007**	[−0.048, 0.062]	> .250	−0.063	[−0.120, −0.006]	.032	−0.084	[−0.229, 0.060]	> .250
Experiment 4	**0.022**	[−0.036, 0.081]	> .250	**0.004**	[−0.057, 0.065]	> .250	**0.040**	[−0.115, 0.196]	> .250
Experiment 5	**0.048**	[−0.009, 0.106]	.102	−0.002	[−0.060, 0.055]	> .250	**0.069**	[−0.075, 0.213]	> .250
Experiment 6	**0.048**	[−0.002, 0.098]	.060	−0.079	[−0.129, −0.029]	.002	−0.046	[−0.164, 0.072]	> .250
Combined results	**0.040**	[0.018, 0.062]	< .001	−0.043	[−0.077, −0.008]	.016	−0.010	[−0.065, 0.044]	> .250

Note: Positive coefficients (highlighted in boldface) indicate results consistent with a given hypothesis. Contrast weights for D and U scores in the native-language (L1) and foreign-language (L2) conditions were as follows—blunted-deontology contrast: {D_{L2}: −3, D_{L1}: +1, U_{L1}: +1, U_{L2}: +1}; heightened-utilitarianism contrast: {D_{L2}: −1, D_{L1}: −1, U_{L1}: −1, U_{L2}: +3}; hybrid-account contrast: {D_{L2}: −1, D_{L1}: 0, U_{L1}: 0, U_{L2}: +1}.

Robustness tests

Given that our experiments varied along several dimensions—participants' native and foreign languages, elicitation format, and dilemma set—we examined whether our findings were moderated by any of these factors. To do this, we partitioned the experiments along a given factor and compared the aggregated coefficient from the blunted-deontology contrast across the experiments that were matched on that dimension with that of experiments that differed on that dimension. For instance, to compare results when participants' native language was German versus English, we aggregated the blunted-deontology contrast coefficients from Experiments 1, 4, and 5 (in which participants' native language was German) and tested whether this coefficient differed reliably from the aggregated coefficient from the blunted-deontology contrast coefficient from Experiments 2 and 6 (in which participants' native language was English). We conducted these analyses using simultaneous estimation equations (Zellner, 1962).

Table 4 provides all 12 pairwise comparisons testing our study permutations. We observed no reliable differences (at the .05 level of significance) in any of these comparisons. Thus, the reduction in deontological responding among foreign-language users appeared to be robust across our various study permutations.

Although elicitation format did not have a statistically significant effect on whether foreign-language use blunted deontological responding (Table 4), we note that the effect of language was least pronounced when the question format explicitly mentioned the negative consequences associated with the utilitarian action (Experiments 3 and 4). This result should be interpreted with caution, but may suggest that using a foreign language makes the aversive features of the choice less salient. Perhaps more noteworthy, however, is the null effect obtained when we compared experiments that crossed native and foreign languages. These comparisons most directly tested whether the reduction in deontological responding among foreign-language users was due to using a foreign language in general rather than using a specific foreign language. Taken together, our results provide evidence that the reduction in deontological responding when participants used a foreign language was robust across a number of contextual factors.

General Discussion

Past research has shown that people are more willing to sacrifice one person to save five when they use their foreign language rather than their native tongue. We investigated why foreign-language use affects moral choice in this way. In particular, we explored whether foreign-language use blunts deontological concerns, heightens utilitarianism concerns, or both. On the one hand, the difficulty of using a foreign language might slow people down and increase deliberation, amplifying utilitarian considerations of maximizing welfare. On the other hand, use of a foreign language might stunt emotional processing, attenuating considerations of deontological rules, such as the prohibition against killing. Across six experiments using different bilingual populations, elicitation formats, and moral dilemmas, we found clear evidence that foreign-language use blunts deontological responding.

We also found no support for the heightened-utilitarianism account. In fact, in three experiments, using a foreign language reliably decreased utilitarian responding. It is unclear why we observed this effect, as the pattern was not consistently elicited by particular study permutations, such as specific language

Table 4. Results From the Robustness Tests

Dimension and comparison	Experimental comparison	$\chi^2(1)$	n	p
Native language				
English vs. German	2, 6 vs. 1, 4, 5	0.07	1,082	> .250
English vs. Spanish	2, 6 vs. 3	1.60	643	.210
Spanish vs. German	3 vs. 1, 4, 5	1.18	829	> .250
Foreign language				
English vs. German	1, 3, 4, 5 vs. 6	0.22	1,035	> .250
English vs. Spanish	1, 3, 4, 5 vs. 2	0.30	1,071	> .250
Spanish vs. German	2 vs. 6	0.00	448	> .250
Crossed languages				
L1: English, L2: Spanish vs. L1: Spanish, L2: English	2 vs. 3	1.31	437	> .250
L1: English, L2: German vs. L1: German, L2: English	6 vs. 1, 4, 5	0.03	840	> .250
Elicitation format				
Judgment vs. judgment with consequences highlighted	1, 2 vs. 3, 4	1.89	862	.169
Judgment vs. choice	1, 2 vs. 5, 6	0.03	871	> .250
Judgment with consequences highlighted vs. choice	3, 4 vs. 5, 6	1.39	821	.238
Dilemma set				
Original vs. revised	1, 2 vs. 3, 4, 5, 6	0.81	1,277	> .250

Note: This table presents the results of tests using simultaneous estimation equations (Zellner, 1962) to compare aggregated coefficients from the blunted-deontology contrasts between experiments that differed in methodology.
L1 = native language; L2 = foreign language.

populations or stimulus materials. One possibility is that decreased utilitarianism among foreign-language users resulted from an increase in cognitive load. Using a foreign language can be cognitively demanding, especially for speakers who are not highly proficient (Plass, Chun, Mayer, & Leutner, 2003), and cognitively demanding tasks impair utilitarian responding (Greene et al., 2008). So, although using a foreign language may have led to lower *D* scores by stunting System 1 processing, it may have also induced lower *U* scores by increasing cognitive load among participants who were not highly proficient in their foreign language (we note that this explanation is independent of the mechanisms we have explored in this article). Because participants rated their foreign-language proficiency at the end of each experiment, we were able to test this explanation. We found tentative support for this explanation in two of the three experiments in which using a foreign language reduced *U* scores. In those experiments, lower levels of foreign-language proficiency were associated with larger reductions in utilitarian responding across language conditions (full details are provided in the Supplemental Material).[6] Although these findings cannot explain why foreign-language proficiency affects *U* scores in some populations and not others, they suggest that cognitive load may play an independent role in the MFLE and should be examined in future research.

Another surprising finding is that the MFLE was not replicated when we examined responses to the kind of moral dilemmas traditionally used to measure utilitarian responding. Our analysis of responses to incongruent moral dilemmas, which directly pitted utilitarian and deontological concerns against each other, showed that foreign-language users were no more willing than native-language users to endorse sacrificial harm. Although this result appears to be at odds with prior research, our stimuli were different, and participants in our study considered multiple dilemmas (rather than a single dilemma) that involved direct contrasts between congruent and incongruent versions. Therefore, multiple variables could account for this discrepancy.

The use of a foreign language affects not just moral choice but decision making more broadly (for a review, see Hayakawa, Costa, Foucart, & Keysar, 2016). To the extent that our findings generalize beyond moral dilemmas, they generate further predictions about the boundary conditions for the effect of foreign-language use on decision making more broadly. For example, Stanovich and West (2008) distinguished decision-making biases that are correlated with cognitive ability from those that are unrelated to cognitive ability. Biases correlated with cognitive ability, such as hindsight bias, outcome bias, and belief bias for syllogistic reasoning, likely reflect System 2 processing, whereas those not correlated with

cognitive ability, such as sunk-cost effects, conjunction fallacies, and anchoring and adjustment, may reflect System 1 processing. This raises the interesting possibility that foreign-language use attenuates decision biases that are associated with System 1 but not with System 2.

The dilemma of whether to sacrifice one life in order to save five is consequential. On the one hand, sacrificing a life is often morally prohibited, and on the other, the lives of five people are surely worth saving. It is surprising that such a fundamental moral choice would be affected by language, and our experiments now provide an explanation. People are more utilitarian when using a foreign language not because they think more, but because they feel less.

Action Editor

Leaf Van Boven served as action editor for this article.

Author Contributions

All the authors contributed to the concept and design of the study. Data collection was performed by S. Hayakawa and J. D. Corey. Data analysis was performed by D. Tannenbaum and S. Hayakawa. All the authors contributed to drafting and editing the manuscript and approved the final version of the manuscript for submission.

Acknowledgments

The authors thank Leigh Burnett, Rendong Cai, and Liz Tenney for commenting on the manuscript and Paul Conway and his colleagues for generously providing the moral-dilemma stimuli.

Declaration of Conflicting Interests

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

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617720944

Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/s5pv8/ and https://osf.io/afdx5/, respectively. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617720944. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Notes

1. We use the terms *deontological* and *utilitarian* responding to refer to responses that are *characteristically* deontological and utilitarian (Greene, 2008; Kahane, Everett, Earp, Farias, & Savulescu, 2015), in that we examine responses consistent with deontological and utilitarian prescriptions, respectively. We do not use these terms as descriptions of participants' meta-ethical beliefs.
2. Across our six experiments, 40 participants (approximately 3% of the sample) had negative U scores. These participants were included in our analyses.
3. Across our six experiments, 7 participants (less than 1% of the sample) had a U score of 1. Because we could not calculate a corresponding D score for these participants (as this would require dividing by 0), we dropped them from our analyses, as recommended by Friesdorf, Conway, and Gawronski (2015).
4. We estimated between-study variance in treatment effects using the standard approach recommended by DerSimonian and Laird (1986).
5. For this analysis, we had two observations per participant (i.e., U and D scores), so we implemented robust clustered standard errors to account for potential nonindependence in observations within participants.
6. Another potential explanation is that the cognitive load of using a foreign language led to more random responding, resulting in both lower U and lower D scores. However, in none of our experiments did we find that lower language-proficiency scores were correlated with a reduction in D scores, which suggests that cognitive load is not likely to account for the decrease in deontological responding among foreign-language users. Therefore, it appears that cognitive load may have selectively interfered with utilitarian responding.

References

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2014). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*, 265–284.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*, 185–216.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.

Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, *29*, 23–40.

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.

Conway, P., & Rosas, A. (2017). *Distinguishing between harm avoidance and outcome maximization in a new battery of process dissociation dilemmas, accounting for guilty victims, fated victims, and selfishness.* Manuscript in preparation, Department of Psychology, Florida State University.

Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1109–1128.

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PLoS ONE, 9*(4), Article e94842. doi: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.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177–188.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Foot, P. (1978). *The problem of abortion and the doctrine of the double effect in virtues and vices.* Oxford, England: Basil Blackwell.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25–42.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin, 41*, 696–713.

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2*, 45–67.

Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS ONE, 10*(7), Article e0131529. doi:10.1371/journal.pone.0131529

Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology, 59*, 8–17.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–80). Cambridge, MA: MIT Press.

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.

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.

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences, 20*, 791–793.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

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.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kant, I. (1959). *Foundations of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis, IN: Bobbs-Merrill. (Original work published 1785)

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.

Klesse, A.-K., Levav, J., & Goukens, C. (2015). The effect of preference expression modality on self-control. *Journal of Consumer Research, 42*, 535–550.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Applied Social Research Methods, Vol. 49). Thousand Oaks, CA: Sage.

Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). New York, NY: Oxford University Press. (Original work published 1861)

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12*, 237–241.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221–243.

Puntoni, S., de Langhe, B., & van Osselaer, S. M. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of Consumer Research, 35*, 1012–1025.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance.* New York, NY: Cambridge University Press.

Schwarz, N. (2011). Feelings-as-information theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 289–308). Thousand Oaks, CA: Sage.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences, 23*, 645–665.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672–695.

Thomson, J. J. (1985). Double effect, triple effect and the trolley problem: Squaring the circle in looping cases. *Yale Law Journal, 94*, 1395–1415.

Wang, Y., Highhouse, S., Lake, C. J., Petersen, N. L., & Rada, T. B. (2017). Meta-analytic investigations of the relation between intuition and analysis. *Journal of Behavioral Decision Making, 30*, 15–25.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57*, 348–368.