

## The role of intentions and outcomes in the foreign language effect on moral judgements

Quarterly Journal of Experimental Psychology  
1–10  
© Experimental Psychology Society 2017  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1747021817738409  
qjep@sagepub.com



Albert Costa<sup>1,2</sup>, Joanna D Corey<sup>1</sup>, Sayuri Hayakawa<sup>3</sup>,  
Melina Aparici<sup>4</sup>, Marc-Lluís Vives<sup>1</sup> and Boaz Keysar<sup>3</sup>

### Abstract

We explore the origin of the foreign language effect on moral judgements by assessing whether language context alters the weight given to intentions and outcomes during moral judgement. Specifically, we investigated whether foreign language contexts, compared with native ones, may lead people to focus more on the outcomes of an action and less on the intentions behind it. We report two studies in which participants read scenarios in which the actor's intentions and the resulting consequences were manipulated. As previously shown, people considered both the actor's intentions and the action's outcomes when assessing the damage, cause, moral wrongness, responsibility, and punishment deserved. However, although the foreign language context reduced the impact of intentions on damage assessment, the overall effect of intention and outcomes on these variables was mainly the same in the foreign and the native language contexts. We conclude that differential weighting of intentions and outcomes is unlikely to account for the impact of foreign language use on moral judgement.

### Keywords

Moral judgement; bilingualism; foreign language processing

Received: 21 March 2017; revised: 14 July 2017; accepted: 10 August 2017

When people use a foreign language, they systematically make different choices than when they use their native tongue (e.g., Costa et al., 2014; Keysar, Hayakawa, & An, 2012). The impact of a foreign language is most pronounced in the context of moral judgements. When faced with a moral dilemma, people tend to choose utilitarian options much more often when using a foreign language than their native language (e.g., Ciolletti, McFarlane, & Weissglass, 2016; Corey et al., 2017; Costa et al., 2014; Geipel, Hadjichristidis, & Surian, 2015). Despite the robustness of this finding, its origin is not well understood. Here, we address a potential explanation of the phenomenon, the possibility that language context affects the relative weight of outcomes and intentions when judging the morality of actions.

The foreign language effect on moral choice has been interpreted in the context of dual process theories of decision-making, which presume the interplay between processes that are fast, intuitive, and affective (System 1), and processes that are deliberate, controlled, and analytical (System 2; e.g., Kahneman, 2003; Sloman, 1996; Stanovich & West, 2000). In the case of moral choice, System

1 processes are thought to promote more deontological, rule-based moral judgements, whereas System 2 processes are associated with the more deliberate consideration of costs and benefits, such as those that promote utilitarian reasoning (e.g., Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). As such, either a decrease in affective System 1 processing or an increase in deliberative System 2 processing could lead to more utilitarian responses.

<sup>1</sup>Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, Barcelona, Spain

<sup>2</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

<sup>3</sup>Department of Psychology, The University of Chicago, Chicago, IL, USA

<sup>4</sup>Departament de Psicologia Bàsica, Evolutiva i de l'Educació, Universitat Autònoma de Barcelona, Barcelona, Spain

### Corresponding author:

Albert Costa, Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, Room 24.329, Edifici Merce Rodoreda, Carrer de Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain.  
Email: costalbert@gmail.com

Explanations of the foreign language effect on moral judgement have appealed to this dual process theory. People are more utilitarian in a foreign language due to a reduction in System 1 processing, an increase in System 2 processing, or both. According to an *increased deliberation account*, the increase in utilitarian responses comes about because using a foreign language promotes more analytic System 2 processes (for a discussion of the effects of this type of processing, see Epstein, 1994; Kahneman, 2003), increasing the contribution of deliberation, which might favour maximising the greater good (see Greene, 2007). Such an analytical approach to a moral situation may focus people more on the outcomes of the actions to be taken than on the means themselves, which could promote utilitarian choices. According to this explanation, foreign language contexts may prompt such analytic thinking more than native language contexts (Costa et al., 2014). According to a *reduced affect account*<sup>1</sup> (Geipel et al., 2015; Hayakawa, Tannenbaum, Corey, Costa, and Keysar (2017), foreign language contexts would reduce the aversive intuitive response to causing harm, especially when it involves exerting physical force on others (e.g., see Cushman, 2013; Miller & Cushman, 2013; Miller, Hannikainen, & Cushman, 2014). When presented with moral dilemmas, individuals often have to decide whether to take such aversive actions to bring about the utilitarian outcome that maximises the greatest good. Hence, to the extent that a foreign language context reduces the automatic elicitation of these aversive responses, one should expect more utilitarian responses compared with when people are in a native language context.

There is ample evidence that when judging others' actions, people consider both the intention of the actor and the consequence of the action (Cushman, 2008; Cushman & Young, 2011). For example, when judging the moral wrongness of a CEO whose workers lost their jobs due to a failed business plan, people will account for whether the CEO's intentions were positive (e.g., a failed attempt to make the company and the workers better off) or negative (e.g., having an excuse to fire some workers). Also, the severity of the judgements varies according to the consequences of the action regardless of the intention. A CEO who caused an entire division of the company to be shut down will likely be judged more harshly than one who caused a single worker to be laid off. Hence, people's judgements depend on both intentions and outcomes.

Both of the accounts presented above would predict that language context may alter the relative weight of intentions and outcomes on people's choices. If foreign language use prompts more deliberate thinking (System 2), or decreases affective reasoning (System 1), or both, this may cause people's judgements to be more affected by the outcomes of a given action assuming that deliberation promotes a focus on the greater good. This is not to say that intentions would not matter in the context of a foreign

language, but rather that they would matter less than in the context of a native language. This is the hypothesis we will explore in the two studies presented here.

There is some evidence suggesting that this hypothesis might be correct. Geipel, Hadjichristidis, and Surian (2016) presented participants with morally relevant scenarios that differed in the intentionality of the actors and the outcome of the actions, and they asked them to rate the actor's moral goodness. In their first study, participants were presented with scenarios in which the actors had dubious intentions, but the outcomes were positive. For example,

*A young couple discovers they are infertile. They decide to adopt a child and successfully pass the exams of the national adoption agency. They are informed that the children that are available for adoption have various birth defects, which most likely caused their biological parents to abandon them. Adopters receive child's pension as well as a disability pension because of the children's condition. The couple does not have money for international adoption. They decide to proceed with the adoption.*

The results were clear: Participants rated the moral goodness of the actors as higher when the texts were presented in the foreign language (German) than when they were presented in the native language (Italian). This is consistent with the idea that a foreign language makes people focus more on the consequences of the actions, thus giving less weight to the actors' intentions. In the second experiment, the actors had positive intentions, but the outcomes of their actions were negative. For example,

*Cristiano deliberately and intentionally gave a homeless man his only jacket, even though it was freezing outside. One hour later two guys saw the homeless person with Cristiano's jacket and beat him up as they thought that he had stolen the jacket.*

Here again, ratings were affected by language context. Moral goodness ratings were higher when the scenarios were in the native language than when they were in the foreign language. This is consistent with the idea that people prioritise outcomes when judging the moral goodness of the situation more in a foreign language than in a native language context.

These results suggest that foreign language use changes the relative weight of intentions and outcomes when making moral judgements. In a foreign language context, outcomes seem to play a more important role, and consequently, intentions matter less. We report two studies where we aimed to replicate these findings and to extend them to other evaluations.

In Study 1, we focus on the roles of intentions in people's judgements. We present two versions of a scenario with the same negative outcome. In one version, the actor responsible for the negative outcome had good intentions,

**Table 1.** Number of participants in each of the eight conditions in Study 2.

		High consequence	Low consequence
Native German	Intentional	85	78
	Accidental	81	86
Foreign English	Intentional	84	81
	Accidental	87	85

and in the other version, he had bad intentions. In general, people give weight to the intention of the actor in judging the morality of an action (e.g., Moran et al., 2011; Young & Saxe, 2009). Therefore, we expect the action to be judged more positively when the actor's intention was positive than negative. The critical issue is whether the magnitude of this difference will depend on the language context of the scenario. If indeed a foreign language context reduces the weight given to the intentionality of the action, then this difference should be smaller when individuals are using a foreign language relative to a native one.

In Study 2, we assess the impact of both intentions and consequences on people's judgements. We do so by presenting scenarios with negative consequences that vary in the degree of harm, either low or high consequences. These consequences resulted from either intentional or accidental actions. This design allows us to assess two issues simultaneously: (a) the importance of outcomes when judging the moral wrongness of an action, and (b) the extent to which such judgements are affected by whether the actions are intentional.

Crucially, we are interested in whether the roles of intention and outcome would depend on the language context. If a foreign language context increases the weight given to the outcomes of an action, then we expect that judgements will be more affected by the difference between low and high consequences in a foreign language than in a native tongue. If a foreign language context reduces the weight given to the intentionality of the actor, then the difference in judgement between intentional and accidental actions will be smaller for those using their foreign language.

### Study 1: bad consequences from good and bad intentions

The goal of this experiment was to evaluate whether a foreign language context reduces the weight people give to intentions when judging the morality of an act. We presented participants with a scenario adapted from Ames and Fiske (2013). Participants read about the boss of a company who made a bad investment decision, causing his employees financial hardship. We manipulated whether the boss had good or bad intentions when he caused the negative outcome. Most importantly, we manipulated whether the scenario was evaluated in a foreign language or a native tongue. While we expect that people would judge the boss

more positively when he had good intentions than bad ones, if a foreign language reduces the weight of intentions, then this difference should be smaller when people use a foreign language compared with a native one.

### Participants

To control for cultural or linguistic factors, we fully crossed the native and foreign languages. One sample of participants had Spanish as the native language and English as foreign, and the other sample had English as the native language and Spanish as foreign. The first sample was collected in classrooms in Catalonia, Spain, with students from a variety of majors such as engineering, psychology, linguistics, and education. We did not sample from foreign language classes. The second sample was collected online in the United States. Half of the sample was recruited by Qualtrics, and the other half were students at the University of Chicago. Within each sample, participants were randomly assigned to perform the task either in their native language or in the foreign tongue. In the Spanish sample, 175 completed the task in Spanish and 195 in English. In the U.S. sample, 181 completed the task in English and 157 in Spanish. For the same information regarding Study 2, see Table 1.

**Demographics.** In all conditions, participants reported their age, gender, and native language. Participants in the foreign language conditions also reported the age of foreign language acquisition, the number of months spent abroad in a country where the foreign language is spoken, and self-rated proficiency for reading, writing, speaking, and listening to the foreign language on a 7-point scale with 7 indicating *full proficiency*. We averaged these four ratings to create an index of overall foreign language proficiency. A summary of these demographic variables is presented in Table 2.

**Comprehension.** After completing the experiment, participants in the foreign language conditions were asked to rate the percentage of the foreign language text that they understood on a scale from 0% to 100%. Participants in the U.S. sample were additionally asked to briefly describe the scenario in English. This extra precaution was taken for the U.S. sample on account of it being an online population, giving us less control over screening out participants who lacked sufficient proficiency.

**Table 2.** Participants' demographic information for all studies.

	Population	Age	% Female	AOA	Months abroad	Foreign proficiency
Study 1	Native Spanish (N=370)	21.92 (4.78)	80.27%	6.45 (2.50)	1.61 (3.69)	5.16 (1.03)
	Native English (N=338)	28.45 (12.49)	73.37%	14.04 (4.76)	9.33 (31.53)	5.29 (0.85)
Study 2	Native German (N=667)	38.80 (14.17)	53.19%	11.52 (4.35)	4.25 (11.81)	5.07 (1.02)

All participants reported age and gender. Participants in the foreign language conditions additionally reported age of foreign language acquisition (AOA), months spent abroad in the foreign language country, and foreign language proficiency (7 = *fully fluent*). Standard deviations are reported in parentheses.

**Table 3.** Number of participants excluded from all studies.

	Population	Non-native	Foreign dominant	Incomplete	Outlier	Comprehension
Study 1	Native Spanish	17	0	3	16	6
	Native English	6	5	0	9	106
Study 2	Native German	0	13	0	14	220

This includes those who were not a native speaker of the target native language, dominant in the target foreign language, had not completed the entire experiment, whose responses were more than 3 standard deviations from the mean, or those who demonstrated insufficient comprehension of the experiment materials.

**Exclusions.** Participants were excluded from the analysis for the following reasons: if they (a) reported that the target native language was not in fact their native language, (b) reported the target foreign language to be dominant to the target native language, (c) failed to complete the experiment, (d) had responses greater than 3 standard deviations (*SDs*) from the mean, or (e) demonstrated a lack of comprehension. Lack of comprehension was determined first by excluding participants who reported understanding less than 50% of the text. For those in the foreign language condition in the Spanish sample, the average self-reported understanding of the text was 88.5% ( $SD=13.4$ ). Only five participants (2.6% of total) reported understanding 50%, and 162 participants (83.1% of total) reported understanding 80% or more (and 74 participants reported understanding 100%; 37.9% of total). For those in the foreign language condition in the U.S. sample, the average self-reported understanding of the text was 85.8% ( $SD=14.5$ ). Only four participants (2.6% of total) reported understanding 50%, and 115 participants (73.3% of total) reported understanding 80% or more (and 35 participants reported understanding 100%; 22.3% of total). Thus, the samples were similar and comparable in terms of foreign language proficiency. In the case of the U.S. sample, we additionally ensured comprehension among the remaining participants by excluding those who did not sufficiently explain the scenario. This was determined by two native English-speaking researchers, one of whom was a fluent Spanish/English bilingual. All exclusions were made prior to data analysis (see Table 3).

### Design and procedure

Participants were randomly assigned to either the native or foreign language condition and were randomly assigned to

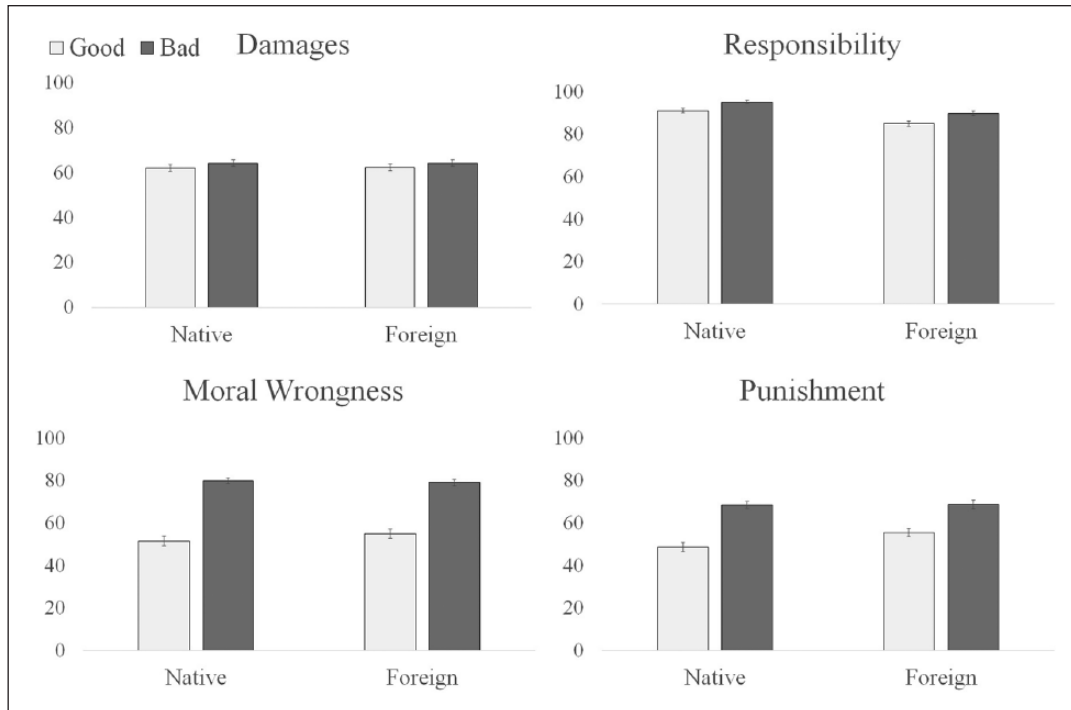
either the "Good Intention" or "Bad Intention" condition. This resulted in a 2 (Language)  $\times$  2 (Intention) between-subject design. Participants were presented with one of the scenarios below (text in italics was common to both conditions):

*Terrance Smith is the boss of a company. His employees get paid by getting a part of the company profits, so if the company makes more money, the employees make more money as well. This has some risk for the employees, but they like it. Sharing the profits also makes the employees work hard. Because the employees work so hard, the company is starting to make a lot of money.*

**[Bad Intentions]** Terrance thought that his employees might become lazy if they continued making a lot of money. So he decided to cause the company to make less money, which he expected would motivate his employees to work harder in the future. That's why Terrance decided to invest in a new project. He was certain that this project would not work and that his employees would therefore make less money than usual. Consequently, he thought this would get his employees to work even harder.

**[Good Intentions]** Terrance thought that his employees would keep working hard if they continued making a lot of money. So he decided to cause the company to make more money, which he expected would motivate his employees to work harder in the future. That's why Terrance decided to invest in a new project. He was certain that this project would work well and that his employees would therefore make more money than usual. Consequently, he thought this would get his employees to work even harder.

*In the end the investment did not work. Nobody in the company (only Terrance) knew why profits were lower than*



**Figure 1.** Mean judgements of damage, responsibility, moral wrongness, and punishment across intention and language. Data are collapsed across the two populations. Error bars represent standard errors.

before. Terrance's own salary did not change, but the employees made less money. This caused them some anxiety. It also made it more difficult to pay bills and take vacations with their families. However, nobody suffered truly terrible financial difficulties or lost their home or car or anything. Terrance did not know about these specific consequences, since he didn't talk to them about personal topics very much.

After they read the scenario, participants responded to the following questions using a 0="None at all" and 100="The most possible" scale. These ratings were the dependent variables:

- How much damage did Terrance's investment in the project cause the employees?
- How much responsibility does Terrance deserve for making the investment?
- How morally wrong was Terrance for making the investment?
- How much should Terrance be punished for making the investment?

If using a foreign language reduces the perceived importance of intentions, we should observe a Language  $\times$  Intention interaction. That is, those using the foreign tongue should be less affected by whether the negative outcome resulted from a good or a bad intention relative to those using their native tongue.<sup>2</sup>

## Results and discussion

To analyse the data, we ran a multivariate analysis of variance (MANOVA) with Damage, Responsibility, Moral Wrongness, and Punishment as dependent variables and Language (native vs foreign), Intention (good vs bad), and Population (native Spanish vs native English) as predictors. We report the analyses examining the raw scores given by participants, but the same pattern of results is found after Z-transforming each score within population. Mean values with standard errors are depicted in Figure 1.

**Damage.** Overall, the degree of perceived harm or damage did not differ depending on Intention ( $F(1, 700)=1.55$ ,  $p=.213$ ,  $\eta_p^2=.002$ ). In addition, there was no main effect of Language ( $F(1, 700)=0.02$ ,  $p=.880$ ,  $\eta_p^2<.001$ ) or an Intention  $\times$  Language interaction ( $F(1, 700)=0.001$ ,  $p=.979$ ,  $\eta_p^2<.001$ ). There was, however, a significant main effect of Population such that the native Spanish speakers perceived greater damage relative to the native English speakers ( $M_s=66.67$  and  $59.50$ , respectively;  $F(1, 700)=24.74$ ,  $p<.001$ ,  $\eta_p^2=.034$ ). More detailed analyses of the effects of population can be found in Supplemental Material.

**Responsibility.** There was a significant main effect of Intention such that the boss with bad intentions was perceived as being more responsible for the negative outcome than the boss with good intentions ( $M_s=92.60$  and  $87.93$ , respectively;  $F(1, 700)=17.55$ ,  $p<.001$ ,  $\eta_p^2=.024$ ). In addition,

there was a main effect of Language such that those using their native language perceived greater responsibility relative to those using their foreign tongue ( $M_s=93.21$  and  $87.28$ , respectively;  $F(1, 700)=28.09$ ,  $p<.001$ ,  $\eta_p^2=.039$ ). However, there was no Intention  $\times$  Language interaction, suggesting that intentions mattered to the same degree in both languages when judging responsibility  $F(1, 700)=0.001$ ,  $p=.979$ ,  $\eta_p^2<.001$ ). No effects of Population were found.

**Moral wrongness.** There was a significant main effect of Intention such that the boss with bad intentions was perceived as being more morally wrong than the boss with good intentions ( $M_s=79.52$  and  $53.18$ , respectively;  $F(1, 700)=194.49$ ,  $p<.001$ ,  $\eta_p^2=.217$ ). However, no main effect of Language was found ( $F(1, 700)=0.22$ ,  $p=.637$ ,  $\eta_p^2<.001$ ). Most critically, we found no Intention  $\times$  Language interaction, suggesting that intentions mattered to the same degree in both languages when judging moral wrongness ( $F(1, 700)=1.19$ ,  $p=.276$ ,  $\eta_p^2=.002$ ). Some effects of Population that emerged are discussed in more detail in Supplemental Material.

**Punishment.** There was a significant main effect of Intention such that the boss with bad intentions was perceived as being more deserving of punishment than the boss with good intentions ( $M_s=68.59$  and  $52.18$ , respectively;  $F(1, 700)=75.35$ ,  $p<.001$ ,  $\eta_p^2=.097$ ). In addition, there was a marginally significant main effect of Language such that those using their foreign language indicated that the boss deserved greater punishment than those using their native tongue ( $M_s=61.86$  and  $58.89$ , respectively;  $F(1, 700)=3.16$ ,  $p=.076$ ,  $\eta_p^2=.004$ ). Unlike the previous three measures, here we do observe a marginally significant Intention  $\times$  Language interaction ( $F(1, 700)=3.06$ ,  $p=.081$ ,  $\eta_p^2=.004$ ). Although both language groups judged the boss with bad intentions as being more deserving of punishment than the boss with good intentions, this difference was larger for those using their native language relative to their foreign tongue ( $M_{\text{difference}}=19.96$  and  $13.08$ , respectively). Follow-up analyses reveal that the two language groups perceived the boss with bad intentions as deserving comparable levels of punishment ( $F(1, 349)<0.001$ ,  $p=.983$ ), but those using their native language believed the boss with good intentions should receive significantly less punishment relative to those using their foreign tongue ( $F(1, 351)=5.61$ ,  $p=.018$ ). No effects of Population were found.

The results of this study show several phenomena of interest. First, intentionality has an effect on people's ratings. This effect is particularly pronounced when rating the moral wrongness and the punishment deserved by the actor. There was a modest effect for judging responsibility and no effect when judging damage. These results are in accordance with previous studies showing that when assessing moral wrong-doing and punishment, people's judgements are less severe when the actor had good intentions than bad

ones. Hence, our study was able to capture the usual pattern of judgements observed when bad outcomes are paired with either good or bad intentions.

However, the crucial question in our study was whether language context affects the tendency to judge actions as more positive when the intention is positive compared with when it is negative. If language makes a difference, this should lead to an interaction between intentionality and language. There is little evidence for this with one exception. When evaluating punishment, the difference between negative and positive intentions was marginally larger in the native language (20%) than in the foreign language (13%). This is consistent with the findings by Geipel et al. (2016) that in a foreign language, intentions receive less weight than in a native tongue.

Together, the results of this study do not show a consistent effect of language context on people's judgements. Given that these results seem to contrast with previous findings about the differential contribution of intentions and outcomes in native and foreign language contexts, we tested the issue further in Study 2.

## Study 2: the contributions of intentionality and consequences

Previous experiments assessed the impact of language separately for intentions and consequences. Study 2 simultaneously manipulated both intention and outcomes. This was done to test not only the hypothesis that using a foreign language reduces the importance of intentions but also the hypothesis that using a foreign tongue should increase the importance of outcomes. Participants were randomly assigned to either their native or foreign tongue and then were presented with one of four scenarios that involved a negative event that varied in terms of both the severity of harm and the intention of the actor.

### Participants

Participants were native German speakers and the data were collected online. The demographic information collected was the same as in Study 1. In addition, we collected foreign language background information from participants in the foreign language condition. The demographic summary is displayed in Table 2, and exclusions are summarised in Table 3. The same exclusion criteria were used as in Study 1, plus participants were excluded if they failed to correctly answer two comprehension questions assessing whether they understood both the intention of the actor and the severity of the consequence.

### Design and procedure

The design was a 2 (Language: German vs English)  $\times$  2 (Intention: Intentional vs Accidental)  $\times$  2 (Consequence:

High vs Low harm) between-subject design. Participants were randomly assigned to one of the resulting eight cells (see Table 1). We used the following scenario about two co-workers who were fighting. The scenarios always started with the following common background story:

*Alexander and Robert work together. Robert often drives Alexander home after work. One day they have an argument as they are getting into Robert's car. While they are still in the parking lot, they start screaming at each other. Eventually Alexander gets tired of fighting, and exits the car. He starts to walk away with his back to the car.*

Then, the story continued in one of these four ways:

- *Intentional—High harm:* Robert sees Alexander behind his car. So Robert puts his car in reverse and intentionally hits Alexander. Alexander is injured and spends 6 months in the hospital.
- *Intentional—Low Harm:* Robert sees Alexander behind his car. So Robert puts his car in reverse and intentionally hits Alexander. Alexander is injured and spends 6 hours in the hospital.
- *Accidental—High Harm:* Robert doesn't see Alexander behind his car. So Robert puts his car in reverse and accidentally hits Alexander. Alexander is injured and spends 6 months in the hospital.
- *Accidental—Low Harm:* Robert doesn't see Alexander behind his car. So Robert puts his car in reverse and accidentally hits Alexander. Alexander is injured and spends 6 hours in the hospital.

After reading the scenario, participants answered the same questions as in Study 1, with the exception of the Responsibility question. This was done to reduce redundancy as the responsibility and moral wrongness questions were very similar to each other. As such, participants judged Damage, Moral Wrongness, and Punishment. As in Study 1, participants filled out a demographic questionnaire and comprehension check following the experiment.

If using a foreign language makes individuals more focused on the outcomes rather than on the intentions, we should observe that those using a native language more clearly differentiate between the intentional and accidental harms, whereas those using the foreign tongue should more clearly differentiate between the low and high severity consequences.

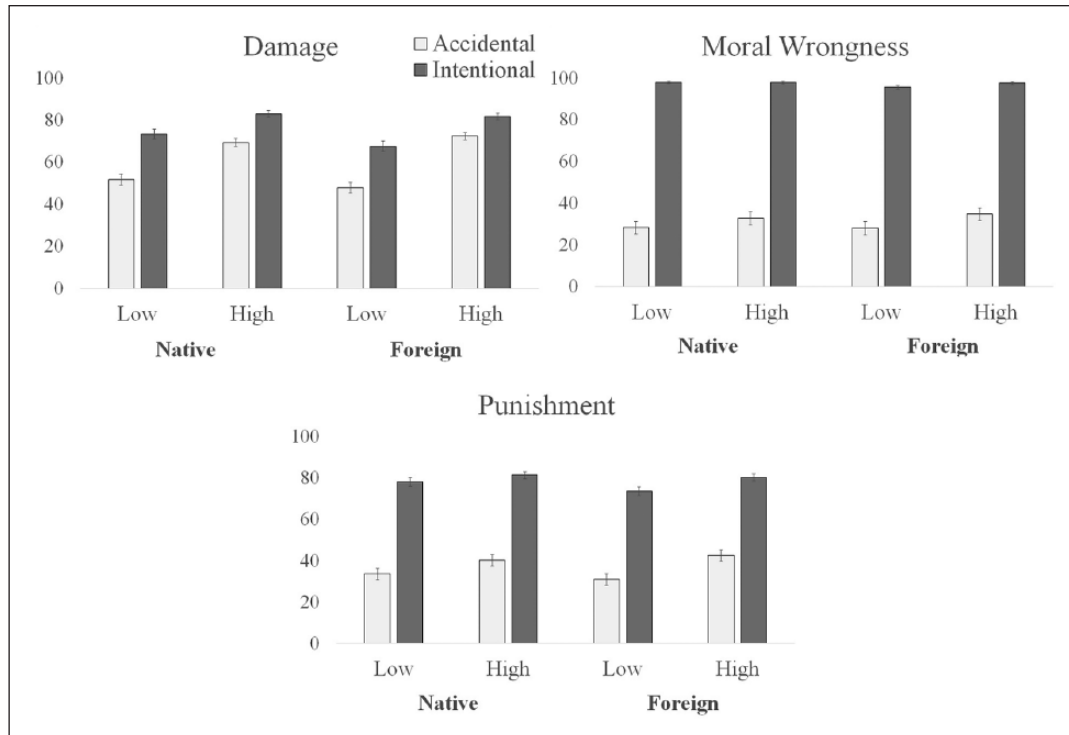
## Results and discussion

To analyse the data, we ran a MANOVA with Damage, Moral Wrongness, and Punishment as dependent variables and Language (native vs foreign), Intention (accidental vs intentional), and Consequence (high vs low) as predictors. Mean values and standard errors are depicted in Figure 2.

**Damage.** Contrary to Study 1, but in line with Ames and Fiske's (2013) findings, participants judged the damage done to be greater when the action was intentional rather than accidental ( $M_s=76.41$  and  $60.26$ , respectively;  $F(1, 659)=115.343, p<.001, \eta_p^2=.149$ ). Participants also perceived the damage to be greater when there were more severe consequences than less ( $M_s=76.60$  and  $60.09$ , respectively;  $F(1, 659)=120.32, p<.001, \eta_p^2=.154$ ), which shows they were paying attention to the details. There was a significant Intention  $\times$  Consequence interaction such that the severity of consequences mattered more when the action was accidental than intentional (High-Low= $21.14$  and  $11.85$ , respectively;  $F(1, 659)=9.534, p=.002, \eta_p^2=.014$ ). As in Study 1, we found no main effect of Language ( $F(1, 659)=1.769, p=.184, \eta_p^2=.003$ ) or an Intention  $\times$  Language interaction ( $F(1, 659)=1.056, p=.304, \eta_p^2=.002$ ). As predicted, there was a marginally significant Consequence  $\times$  Language interaction such that the severity of consequences mattered more when using the foreign language relative to the native tongue (High-Low= $19.36$  and  $13.63$ , respectively;  $F(1, 659)=3.625, p=.057, \eta_p^2=.005$ ). Finally, we found no evidence of an Intention  $\times$  Consequence  $\times$  Language interaction ( $F(1, 659)=0.161, p=.688, \eta_p^2=.001$ ).

**Moral wrongness.** As in Study 1, participants judged intentional bad outcomes as more morally wrong than accidental ones ( $M_s=97.22$  and  $30.95$ , respectively;  $F(1, 659)=1,629.954, p<.001, \eta_p^2=.712$ ). Participants also perceived the scenarios with more severe consequences to be more morally wrong relative to those with less severe consequences ( $M_s=65.73$  and  $62.44$ , respectively;  $F(1, 659)=4.021, p=.045, \eta_p^2=.154$ ). However, unlike ratings of Damage, there was no Intention  $\times$  Consequence interaction ( $F(1, 659)=1.921, p=.116, \eta_p^2=.003$ ). As in Study 1, we found no main effect of Language ( $F(1, 659)=0.019, p=.892, \eta_p^2<.001$ ), nor did Language interact with Intention ( $F(1, 659)=0.478, p=.490, \eta_p^2=.001$ ) or Consequence ( $F(1, 659)=0.455, p=.500, \eta_p^2=.001$ ). Finally, we found no evidence of an Intention  $\times$  Consequence  $\times$  Language interaction ( $F(1, 659)<0.001, p=.982, \eta_p^2<.001$ ).

**Punishment.** As in Study 1 where the boss with bad intentions was perceived as more deserving of punishment than the boss with good intentions, there was a significant main effect of Intention such that intentionally physically harming someone resulted in higher punishment than accidentally doing so ( $M_s=78.31$  and  $36.79$ , respectively;  $F(1, 659)=583.647, p<.001, \eta_p^2=.470$ ). Participants also perceived the scenarios with more severe consequences as more punishable relative to those with less severe consequences ( $M_s=61.08$  and  $54.02$ , respectively;  $F(1, 659)=16.873, p<.001, \eta_p^2=.025$ ). However, there was no Intention  $\times$  Consequence interaction ( $F(1, 659)=1.424, p=.233, \eta_p^2=.002$ ). Unlike Study 1, there



**Figure 2.** Mean judgements of damage, moral wrongness, and punishment across intention (accidental vs intentional), consequence (low vs high harm), and language (native vs foreign). Error bars represent standard errors.

was no main effect of Language ( $F(1, 659)=0.732$ ,  $p=.393$ ,  $\eta_p^2=.001$ ), nor was there an interaction between Language and Intention ( $F(1, 659)=0.640$ ,  $p=.242$ ,  $\eta_p^2=.001$ ) or between Language and Consequence ( $F(1, 659)=1.439$ ,  $p=.231$ ,  $\eta_p^2=.002$ ). Finally, we found no evidence of an Intention  $\times$  Consequence  $\times$  Language interaction ( $F(1, 659)=0.068$ ,  $p=.794$ ,  $\eta_p^2<.001$ ).

This study reveals several interesting observations. First, the actor's intention affected all judgements of moral wrongness, damaged caused, and punishment deserved, with lower values being assigned when the harm was caused accidentally, consistent with Ames and Fiske (2013). Also, as expected and in accordance with Study 1, this effect of intention is especially large for judgements of moral wrongness and punishment. Second, the severity of the consequences also affects participant's judgements, with more extreme ratings assigned for more severe consequences. This is especially so for the judgements about damage and punishment. Although the severity of consequence did have a significant effect on judgements of moral wrongness, the effect was relatively small, suggesting that intention is more important than outcome for determining moral wrongness, consistent with the previous literature (e.g., Cushman, 2008). Together, these results reveal that our study was able to capture the contribution of intentions and the severity of outcomes on people's judgements. Furthermore, as expected, moral wrongness and punishment judgements were more affected by intentionality than by the severity of

the outcome, but perceived damage was affected by intentionality and the severity of the outcome to a similar extent.

Thus, having shown the sensitivity of the design to capture the contributions of intentions and outcomes to participants' ratings, the question is whether such contributions depends to some extent on the language context. We found only partial evidence for this.

The overall effect of language on people's ratings was similar for native and foreign language contexts. That is, it is not the case that language affects the perceived amount of moral wrongness, damaged caused, or deserved punishment. This result replicates the observations of Study 1. More important for the main hypothesis is that language context does not affect the way intentionality is treated when judging people's actions. Finally, language context has little effect on how consequences affect participant's ratings. The one exception was for judgements of damage, which were more affected by severity of consequences when described in the foreign language relative to the native tongue.

## General discussion

We presented two studies that tested a potential explanation of the foreign language effect on moral judgements: that language context may affect the importance of intentionality and outcomes when people judge morally relevant acts. We considered the possibility that, compared with using a native language, using a foreign one may make people's



judgements less affected by the intentions of the actor and more affected by the outcomes of the action.

The results of our two studies cast doubt on the validity of this hypothesis as we found little evidence that language contexts affect people's judgements on the moral wrongness of the action, the responsibility of the agent, the damage caused, or the punishment deserved. Furthermore, the impact of intentions and consequences seems to be similar across languages for the most part.

These results seem inconsistent with findings reported by Geipel et al. (2016) regarding how language context alters the impact of intentions and outcomes on people's judgements. However, a direct comparison between the two studies is difficult because we did not attempt an exact replication of Geipel et al.'s studies. There are differences in the languages used here and in their study, and the conditions we used are not fully comparable to their conditions. In their study, participants were asked about the moral goodness of the actor, whereas we ask about damage, responsibility, moral wrongness, and punishment. Asking about goodness or wrongness may focus participants on different aspects of the scenario, giving more or less weight to the intentions versus outcomes. The negative frame of asking about "wrongness" of an action may increase the focus on the intention compared with the positive frame on the "goodness" of an action (e.g., Pizarro, Uhlmann, & Salovey, 2003). This may reduce the focus on the outcomes. Asking people about the goodness of an action may, in turn, lead people to give more consideration to the outcome of the action. It is unclear how this change in the valence of the question may affect the impact of a foreign language. However, it may be the case that the effect of the foreign language only emerges when there is sufficient consideration of outcomes overall. Given that Geipel et al. (2016) did not vary the outcome within an experiment, we are unable to compare the consideration of outcomes across studies. It will be interesting to see whether a fully crossed design using materials similar to Geipel et al.'s would still show differential effects of consequences and intentions as a function of language context.

Another reason for the discrepancy between our findings and Geipel et al.'s (2016) might relate to the presence of trade-offs. The foreign language effect has mainly been found when there is a trade-off between the negative and positive outcomes of a choice. This is certainly the case for moral dilemmas in which sacrificing is required to increase the greater good, potentially to a certain extent for moral transgressions where breaking an abstract moral rule is required to satisfy some selfish need or desire and arguably for situations where intentions and outcomes do not match. In general, people avoid trade-offs when they are emotionally aversive (e.g., Luce, Payne, & Bettman, 1999). Therefore, it may be that language effects emerge, in part, due to a reduction in the emotional aversion

elicited by such trade-offs when these are highly salient. It may be the case that no such trade-offs exist for the scenarios we used or at least not for those used in Study 2. Granted, such an explanation is mere conjecture at this point and would need to be empirically tested.

Although overall our results suggest that language context does not as a rule affect the importance of intentions and consequences in people's judgements, there is some evidence to support this idea. In both studies, we found a similar result regarding the role of language in how people determine deserved punishment. In Study 1, people thought that the boss deserved more punishment when his intention was bad than when it was good, but this difference was smaller in the foreign language than in the native language. In Study 2, we observed a similar pattern, as the difference between punishment ratings in accidental versus intentional harms was a bit smaller in the foreign (40 points) than in the native language (43 points). Indeed, when combining the two studies, the interaction between language and intentionality is significant ( $F(1, 1,369)=4.42, p=.036, \eta_p^2=.003$ ). This outcome is consistent with Geipel et al.'s (2016) results. Although this should be interpreted with caution, it may have important implications for society, and especially for justice officials when forming public juries. It therefore might deserve further exploration and confirmation.

A potential caveat to these studies is that they used different types of samples and procedures (online vs collected in classrooms), which may be problematic for comparing the results across studies. However, notice that the foreign language effect was absent for all samples and that there were very few differences between samples on the measures taken. Hence, it seems that the lack of a foreign language effect cannot be attributed to differences between the samples.

The goal of the current studies was to shed light on the origin of the foreign language effect on moral judgements. In the Introduction, we discussed two accounts of this phenomenon, one that argues that foreign language use increases deliberation and another that argues that it decreases affect. We further argue that both of the accounts may predict that language context may alter the weight of intentions and outcomes on people's choices and that may be at the basis of an increase in utilitarian choices associated with foreign language contexts compared with native language.

The results of this article do not support this hypothesis. This does not mean that a dual process account is wrong. What the results of the present studies do suggest is that the mechanism by which a foreign language affects moral judgement is one other than altering the weights given to intentions and outcomes when making judgements.

### Acknowledgements

The authors thank Juan Botella for his help with the statistical analysis.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by a grant from the University of Chicago's Wisdom Research Project and the John Templeton Foundation #37775 and a grant by the National Science Foundation #1520074 to the University of Chicago. This research was partially funded by grants from the Spanish Government (PSI2011-23033, Consolider Ingenio 2010 CSD2007-00048, and Ministry of Economy and Competitiveness PSI2014-52181-P), from the Catalan Government (SGR 2009-1521), and from the European Research Council under the European Community's Seventh Framework (FP7/2007-2013 Cooperation grant agreement 613465-AThEME). Joanna D Corey was supported by a grant from the Catalan Government (FI-DGR). Marc-Lluís Vives was supported by a grant from the Spanish Government (BES-2015-071581). Sayuri Hayakawa was supported by a Harper dissertation fellowship from the University of Chicago.

## Supplemental material

Supplementary Material is available at: [journals.sagepub.com/doi/suppl/10.1177/1747021817738409](http://journals.sagepub.com/doi/suppl/10.1177/1747021817738409).

## Notes

1. Elsewhere, this has been referred to as the "reduced intuition account." Both terms refer to a reduction in so-called System 1 processes, which are automatic, intuitive, and affective.
2. As suggested by an anonymous reviewer, it is possible that the intentions in Study 1 were not seen as different given that goal was the same (motivating the employees to work hard and earn good money). However, this seems highly unlikely given that we find a robust effect of intention for most of the measures in this study. If the good and bad intentions were perceived similarly by the participants, then we should not have found an effect of intentionality (especially not for the measure of moral wrongness, which is largely driven by intentionality).

## References

- Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*, 1755–1762.
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, *29*, 23–40.
- 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), e94842.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353–380.
- Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and Social Psychology Review*, *17*, 273–292.
- Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from non moral psychological representations. *Cognitive Science*, *35*, 1052–1075.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, *59*, 8–17.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition*, *154*, 34–39.
- Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Update*, *11*, 322–323.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400. doi:10.1016/j.neuron.2004.09.027
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108. doi:10.1126/science.1062872
- Hayakawa, S., Tannenbaum, D., Corey, J. D., Costa, A., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign language effect on moral judgment. *Psychological Science*, *28*, 1387–1397. doi:10.1177/0956797617720944
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.
- 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.
- Luce, M. F., Payne, J. W., & Bettman, J. R. (1999). Emotional trade-off difficulty and choice. *Journal of Marketing Research*, *36*, 143–159.
- Miller, R. M., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*, 707–718. doi:10.1111/spc3.12066
- Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, *14*(3), 573–587. doi:10.1037/a0035361
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 2688–2692.
- Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise the role of perceived metadesires. *Psychological Science*, *14*, 267–272.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*, 701–717.
- Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*, 2065–2072.