If only judgments were different in a foreign language

Elias Eckardt

S5252342

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number:10

Supervisor: Dr. Kai Epstude

Second evaluator: Dr. Stacy Donofrio

In collaboration with: Lara Bakker, Anna de Wind, Britt Hollander, Naomi Meijnen, and Rinke Monja.

June 27th, 2025

Abstract

This study investigates whether the Foreign Language Effect influences blame assignment in a counterfactual scenario and if thinking in a foreign language mitigates the halo effect. Centered around the dual-process theory and the reduced affect hypothesis, we hypothesized that processing in a foreign language (English) would decrease heuristic and emotional thinking while promoting analytical processing. In a counterfactual scenario, we expected a decrease in blame assignment and the mitigation of the halo effect, specifically for unrelated traits. A 2x2 between-subjects experiment (Language: native vs. foreign; scenario mutability: mutable vs. immutable) was conducted with 595 native Spanish-speaking participants. Results showed that, contrary to our hypothesis, participants in the foreign language (vs. native) condition assigned significantly more blame. Although blame scores increased in the counterfactual scenarios, this effect was not robust across the non-parametric analyses. In support of the hypothesis, centered around the trait "hard-working," unrelated traits were rated significantly lower in the Foreign Language (vs. native) condition, suggesting a reduction in the halo effect. Traits related to a central trait were rated similarly across both language conditions. The study encountered a limitation in that the scenario failed to induce emotional involvement and lacked moral ambiguity in the attribution of blame. As this serves as a possible explanation for the rejection of our hypotheses, future research could improve on the study design to build on our findings regarding counterfactual blame attribution in a foreign language. The mitigation of the halo effect in a foreign language contributes to the understanding of how the Foreign Language Effect mitigates cognitive biases.

Keywords: blame, foreign language effect, counterfactuals, halo effect, reduced affect, dual-process theory

Does the foreign language effect apply to blame assignments in counterfactuals?

"If the Stasi had violently suppressed the protests of Leipzig in October 1989, the DDR would not have collapsed, the Berlin wall would still stand, and Germany would remain divided today" (ARTEde, 2024). Engaging with this example of an if-only thought revolving around the reunification of East and West Germany enables the reader to attribute the success of the reunion to the demonstrators and assign blame in this downward counterfactual scenario to the Stasi. Such counterfactual thoughts create and simulate an alternative sequence of events where a change in one element can reshape the perspective and judgment of the outcome and actors involved. Imagining what might have been can increase the perceived behavioral control of a responsible actor, making the negative outcome appear more preventable, which can enhance emotional affect and elicit stronger feelings of blame (Alicke et al., 2008).

Our judgment is not just influenced by the way we think; it can differ depending on which language we use. People tend to use more rational and utilitarian moral judgments in a foreign language compared to their native language. As we tend to have a stronger emotional resonance with our thoughts in our native language, thinking in a foreign language reduces emotional affect in our judgment, prompting more rational and analytical thinking (Keysar et al., 2012). If negative emotional affect can be intensified by imagining and simulating a better outcome in a counterfactual scenario, thus increasing blame, processing in a foreign language could mitigate this judgment. Does the Foreign Language Effect reduce blame assignments in a counterfactual scenario?

When we blame people, we do not just base our judgment on observed behavior; we also let our perception of the person influence our judgment. If people are perceived as socially attractive, they face significantly lower assignments of blame (Alicke & Zell, 2009). To appear socially attractive, it is often enough to be described positively on one central trait,

such as being friendly. Such traits prompt an overgeneralized positive image, leading people to expect this person to possess other positive traits, such as loyalty. This heuristic thinking gives rise to the halo effect, which entails overemphasizing the evaluation of a central or salient trait and then applying this evaluation to other related and unrelated character traits (Thorndike, 1920). Given that this cognitive bias is amplified by heuristic and emotional processing, the foreign language effect may reduce overgeneralizations by creating emotional distance and promoting analytical thinking. Does thinking in a foreign language reduce the Halo Effect?

Literature review

The Foreign Language Effect

With its roots in bilingual research, people tend to express negative emotional responses in their native language and utilize their second language to practice self-control or emotionally distance themselves from their emotions (Pavlenko, 2008). According to Keysar et al. (2012), using a foreign language reduces biases such as loss aversion and the framing effect by increasing cognitive load, which decreases the capacity for emotional processing. To process in a foreign language, more analytical and conscious thinking is required, which involves creating psychological and emotional distance and deliberation from the information. Additionally, emotional detachment is evoked due to a reduction in the emotional association with words and events that were initially formed in the native language, making them feel more distant and less immediate in a foreign language. Similarly, psychological distance is created due to linguistic dissociations with the self, resulting in less personal involvement and a more objective stance (Costa et al., 2014).

Accordingly, people are more likely to choose utilitarian options in moral dilemmas, indicating a shift from emotional and intuitive processing to analytical and rational processing (Miozzo et al., 2020). According to the dual-process theory, two processing

systems are differentially activated depending on the language; System 1 is the relatively fast, intuitive process that is largely unconscious, affective, and heuristic, while System 2 is the slower, analytical process relying on conscious, deliberate, and systematic thinking (Barrouillet, 2011). Thinking in a foreign language promotes System 2 processing, through which judgments commonly made by System 1 are more rational and less emotionally charged.

Blame attributions

A decrease in emotional arousal could significantly alter moral judgment, specifically in how we assign blame to people. Blame consists of two basic elements: firstly, identifying a person responsible for the caused harm; secondly, assigning accountability for the harm to that individual (Weiner, 1995). Regarding the decision-making process for assigning blame, according to the dual-process model, people employ heuristic (System 1) and analytical thinking (System 2) to form their judgments.

According to Alicke (2000), the culpable control model proposes three structural links that explain how people engage in rational attributions of blame. The responsible agent is evaluated on the intention of their behavior (Volitional Behavior Control), their causal relationship to the outcome (Causal control), and to what extent the outcome was intended (Volitional Outcome Control). Due to emotional arousal, a shift from rational (System 2) to heuristic thinking (System 1) can be induced. This has been termed "swift blame," which is more susceptible to inaccurate judgment and distorted beliefs that could lead to increased fluctuating assignments of blame. By judging the responsible person based on their character rather than their behavior, blame can be adjusted to their perceived overall image. Additionally, following appraisal theory (Lazarus, 1991), automatic judgments provoke affective arousal, which further shapes the interpretation beyond rational attribution of blame. These cognitive shortcuts can lead to an increase in blame assignment by overestimating the

intention behind the behavior and outcome, and the causal control that is rationally more accurately assessed (Skarlicki et al., 2017).

Blame in a foreign language.

As processing in a foreign language facilitates analytical thinking (System 2) and decreases emotional reactivity, the heuristic thinking that leads to a distortion of moral judgment and, therefore, higher assignments of blame may be mitigated. This effect has been partially demonstrated, as the Foreign Language effect mitigates the effects of the causality bias that commonly leads to increased blame assignments (Díaz-Lago & Matute, 2018). Contrarily, according to Bodig et al. (2019), reduced emotionality may increase the likelihood of assigning blame, as empathy facilitates perspective-taking, a crucial factor in determining responsibility. Accordingly, the foreign language effect could have an adverse impact if the actor's intention and behavioral control are ambiguous, whereby empathy could otherwise increase the perception of situational causes.

Therefore, thinking in a foreign language may only be able to decrease blame if empathy for the actor is irrelevant and the scenario elicits strong emotional affect that amplifies swift blame. One can increase emotional arousal by imagining and simulating a negative outcome that could have been prevented if only something had been different.

Counterfactual thinking

Engaging in an 'if-only' thought allows the mind, in the hindsight of an event, to reconstruct and simulate an alternative outcome. If only a factor that initially seemed to contribute to the chain of events had been different, the outcome would have been different. The 'if-only' scenario of the violent suppression of the demonstrations in Leipzig (1989), which disrupted the reunification of Germany, is an example of a downward counterfactual. It focuses on how the past might have been worse if only violence had been used and can bring out feelings of relief and appreciation, serving an affective function. On the other hand,

upward counterfactuals enable individuals to imagine how the past might have been better than it was; for example, "If only I had studied more for my exam, I would have passed with ease" (Epstude & Roese, 2008). By simulating a better alternative, one is reminded of the negative outcome, causing negative emotions to resurface. This negative arousal, especially, evokes strong feelings of blame and regret towards the factors deemed responsible for the outcome (Alicke et al., 2008). We mentally create internal representations of events and objects that are cognitively structured in a way that helps us understand and adapt to our environment. Counterfactual thoughts build upon this premise by introducing mutability, referring to the extent to which these internal representations appear plausible to be changed and then simulated to the improved outcome (Epstude & Roese, 2017).

According to the simulation heuristic, if it is easy to simulate a counterfactual thought, the outcome is perceived as more preventable, which is accompanied by greater assignment of blame (Kahneman & Tversky, 1982). Blame is increased due to the overestimation of perceived causality and intention of the actor and their actions. Identifying the cause and establishing blame, although closely related, need to be differentiated. In high norm-breaking behavior, blame is high and closely related to the cause; however, in lower norm-breaking behavior, blame is substantially lower and less associated with the cause and responsibility (Spellman & Gilbert, 2014).

Counterfactual blame attribution in a foreign language

In upward counterfactual thoughts, blame can be increased due to the intensified emotional affect evoked by the regret of not having obtained a better outcome and an increased perception of preventability, leading to an overestimation of causality and intention. The Foreign Language Effect could decrease the assignments of blame in counterfactual thinking, as it reduces emotional affect and promotes analytical thinking to mitigate the simulation heuristic, leading to a more rational attribution of blame.

The Halo effect

When people evaluate personal characteristics, there is a tendency to match the evaluation of one trait with a trait that is already established. Thus, based on the positive description of an individual as attractive, people tend to match this with other favorable traits, such as intelligence. A single central trait can prompt heuristic and emotional thinking, leading to overgeneralized evaluations that influence judgments of other character traits and shape the overall perception of a person's characteristics. This has been termed the halo effect by Thorndike (1920).

Concerning blame, evaluations of traits related to the behavior can lead to more negative assessments, as these traits are semantically and logically linked to the wrongdoing. In contrast, unrelated traits are more influenced by the individual's overall image and are only affected if the blame is severe enough to alter that general impression (Murray et al., 2024). The linguistic context carries associative and affective properties that should be considered. As human characteristics are perceived as interconnected, some traits are closely related to each other (e.g., kindness and empathy), while others are more unrelated (e.g., kindness and ambition). The likelihood of two traits being judged as co-occurring in an individual can be predicted mainly by three factors: cosine similarities (the extent to which traits appear in similar contexts), valence (the emotional tone), and word frequency. (Westbury & King, 2024). Therefore, related traits are rated higher, as they are linked to the central trait through semantic meaning and logical associations within the linguistic context. In contrast, unrelated traits are rated lower, primarily influenced by the general impression created by the central trait and affective associations in the linguistic context.

Halo effect in a foreign language

The halo effect arises from heuristic thinking (System 1) and linguistic context, which can be attenuated by processing information in a foreign language, promoting analytical

thinking (System 2) and increasing psychological and emotional distance while weakening affective and contextual associations (Keysar et al., 2012). Related traits depend primarily on logical semantic associations with the central trait; processing in a foreign language could only diminish affective associations established in the native linguistic context, thereby selectively mitigating the halo effect. As for unrelated traits, thinking in a foreign language could reduce the affective influence of the general impression, leading to a more rational evaluation, while disassociating the affective linguistic context for their judgment, possibly leading to a significantly diminished halo effect.

Overview of the study

Based on the presented research on counterfactual thoughts, we assume that being presented with a scenario where the negative outcome was mutable, participants simulate an 'if-only' scenario, which increases their emotional affection, perception of intention, and causality of the actor, and thereby increases their assignment of blame. We hypothesize that Participants processing this scenario in their foreign language will decrease affective heuristic thinking, which can lead to the overestimation of causal control and intention of the actor, and therefore assign less blame than participants in their native language. Lastly, we assume that due to the reduction in emotionality and linguistic dissociation, participants in their foreign language exhibit decreased heuristic thinking, which leads to the overarching evaluation of a central trait over unrelated traits.

Hypothesis 1

Participants in the mutable condition will attribute more blame than participants in the immutable condition.

Hypothesis 2

Participants in the foreign language condition will assign less blame than participants in the native language condition.

Hypothesis 3

The difference in blame assignment between the mutable and immutable conditions will be greater in the native language condition than in the foreign language condition.

Hypothesis 4

Based on the presented central trait, participants in the foreign language condition will rate unrelated traits lower than participants in the native language condition, while related traits will not differ significantly across conditions.

Methods

Participants

The study initially recruited 629 participants primarily from Spain, Argentina, and Mexico through the online platform Prolific. After implementing robust bot detection protocols, 17 bot accounts were identified and removed from the dataset. The final analytical sample comprised 595 valid participants (298 English speakers and 297 Spanish speakers) between 19 and 66 years of age. These participants were evenly distributed across the *Mutable* (n = 314) and *Immutable* (n = 281) experimental conditions. All participants received monetary compensation upon completion. While the study used an educational scenario, participation was open to both students and non-students, as long as they met the age and language criteria. Participants were native Spanish speakers with an intermediate to advanced English proficiency.

Exclusion criteria included never using one of the two languages and being flagged by the control checks. These checks were used to assess whether participants were paying attention or not by asking the participants to choose a specific answer option. Demographic data has been collected, such as age, gender, nationality, country of residence, English proficiency level, and the frequency of English use. The study has received ethical approval

from the ethics board. Participants have been provided with an informed consent sheet. This was to ensure that they were aware of their rights, including the ability to withdraw at any time, and to confirm the confidentiality and anonymity of their response.

Design

This study used a between-subjects experimental design. Participants from a convenience sample either answered the online questionnaire in Spanish as their native language or in English as their foreign language. Upon participating in the study, the participant was presented with a scenario followed by being randomly assigned to one of the two outcomes. Afterwards, the participant was asked to evaluate this scenario on a quantitative questionnaire. This study followed a 2(language) x 2(outcome) design. Language (Spanish or English condition) and outcome (*mutable* or *immutable* condition) are the two independent variables, and the amount of blame attribution is the dependent variable. Additionally, English proficiency and demographic information of the participants have been collected to control for possible confounding variables. This study has been approved by the ethics commission at the University of Groningen in accordance with EC-BSS guidelines for low-risk studies.

Materials and Measures

In this study, we used a two-part survey presented in either the participants' native language (Spanish) or a foreign language (English), which makes up one of the two factors with 2 levels. The survey had an approximate duration of 7 minutes and 29 seconds.

Demographics and Foreign language learning (FLL)

Firstly, information about certain demographics, including gender identity, age, nationality, and country of residency, was collected. In addition, we asked for the parents' native language, hereby focusing on our languages of interest individually: Spanish and English. FLL was assessed by asking multiple questions, including frequency of use in daily

life, self-described proficiency, age of acquisition, and the setting where most of the language skills were learned. The answer options and scales can be found in Appendix A3.

The counterfactual scenario

The scenario was divided into two parts. First, all participants read a general backstory set in a university context about a student, Nora, preparing for an important exam and relying on a tutor, Damian, who was asked for help. He was portrayed as negligent and did not read her email. Secondly, participants were randomly assigned to one of the two different outcomes that were evenly distributed, making up the other factor with 2 levels. In both possible outcome versions, the student failed her exam and was emotional because of the high stakes of the exam. The difference in the randomized scenario options was the increase in the possibility of a different outcome for Nora's situation. This was operationalized by stating that she knows for a fact she would (mutable condition) or would not have passed (immutable condition) with the help of her tutor. Here are the two different conditions. All scenario parts can be found in Appendix A5.

Blame Attribution and Sympathy

Blame attribution and sympathy were measured by six questions about how participants assign blame to Damian as well as Nora and their sympathy towards Nora, as well as help expectancy. An example is: "Damian is to blame for Nora's bad grade". The questions were asked on a Likert scale from 1 ('strongly disagree/not at all') to 5 ('strongly agree/extremely'). For each of the current and following questions, with the exception of the "Halo Effect" scale, the reaction time was measured. The questions of this part can be found under Appendix A7.

Emotionality

Additionally, we measured the extent how emotional/personal the scenarios felt to the participants. This was done on the same scale as the previous part. The question of this part can be found under Appendix A7 (Question 7)

Halo effect

This part starts with a reminder of Nora's attribute from the scenario, described as "hard working", and is followed by eight adjectives, that needed to be rated on a 5-point Likert scale from 1 ('not at all likely') to 5 ('extremely likely') in response to the question whether Nora is likely to possess these traits. These adjectives were differentiated into four traits related to "hard working" averaged into the variable "related traits" (e.g, ambitious), and four unrelated traits to "hard working" averaged into the variable "unrelated traits" (e.g, kind). We also included a plausibility check here to ensure participants perceived the scenario as realistic. This was asked as: "The scenario is plausible/realistic to me," scaled on a 5-point Likert scale from ('not at all') to 5 ('very much'). Response time was not measured here. All details of this part can be found under Appendix A8

Correspondence Bias and Likeability

The next nine statements battery had to be rated on a 5-point Likert scale from 1 ('not at all') to 5 ('very much'), in response to being asked how much they applied to the scenario. The statements centered around different correspondence manipulations about Damian and oneself, like " Damian's failure to help was due to his personality", and likeability, like "I think Damian is a likeable person". We included one control check, and response time was measured again. All details of this part can be found under Appendix A9.

Validity and Reliability

Generally, since the study focused on subjective moral judgments rather than standardized psychological constructs, traditional measures of validity and reliability did not directly apply. For the internal consistency of the blame measure used in the analysis (consisting

of “Damian is at fault for Nora’s bad situation” and “Damian is to blame for Nora’s bad grade”), the correlation between the two items was calculated. The two items showed a significant positive correlation with $r=0.67$, and $p<0.001$, which suggests acceptable item agreement (Eisinga et al., 2012). Concerning the measure of the halo effect, the reliability of the variable “related traits” was high with a *Cronbach’s α=0,86,* and that of the variable “Unrelated traits” was acceptable, with a *Cronbach’s α=0,73*

Procedure

On the online platform Prolific, the participants could either sign up for the Spanish version of the questionnaire (native language condition: L1) or English (foreign language condition: L2). These were the two language conditions. If a participant signed up for the Spanish version, they were not able to sign up for the English version, and vice versa. After signing up, the participants were directed to our questionnaire on Qualtrics (www.Qualtrics.com). They could participate in their own time on a device of their choice. First, they were asked to read the information about the research and then guided to the consent form (Appendix A1 and A2). Only the participants who gave informed consent could continue with the questionnaire. Following the questions of demographic information, they were randomly assigned to either the immutable condition or the mutable condition. The last part of the questionnaire consisted of rating scales that measured moral judgement. This included questions about the amount of assigned blame (to both the tutor and student). These responses were timed. Lastly, a plausibility check was conducted to see if the scenario was believable, which was measured on a 1-5 Likert scale with “completely disagree” as one and “completely agree” as five. At the end of the study, the participants were thanked and given the opportunity to contact the researchers for more information.

Results

Blame Scores

A 2x2 between-subjects ANOVA (Appendix B: Table 3) was conducted with Language (Spanish vs. English) and scenario outcome (Mutable vs Immutable) as factors. Prior to conducting the analyses, the assumptions of normality, homogeneity of variances, and independence were checked. The distribution visible on the histogram (Appendix B: Figure 1) is left-skewed and does not appear normally distributed. To assess further, the Q–Q plot of residuals (Appendix B: Figure 2) showed approximate normality as well as the Shapiro–Wilk test, W =0.97, p=.12. The data does, therefore, moderately violate the assumption of normality. Homogeneity of variances was tested using Levene's Test for Equality of Variances and was found to be non-significant (p>.001), while there was no extreme outlier found in the boxplot (Appendix B: Figure 3).

Hypothesis 1

The main effect of counterfactual thinking on blame assignment was significant, *F(1, 591)=4.93, p=.027.* Participants in the mutable condition assigned higher levels of blame, *M=1.87 (SD=0.99),* to Damian than the participants in the immutable condition, *M=1.61 (SD=0.68).* Due to the violations of the assumptions, an additional non-parametric analysis (*Mann–Whitney U Test*) was conducted, *U=44,728.50, Z=-0.95, p=.344*, yielding non-significant results, contradicting the hypothesis.

Hypothesis 2

A significant main effect emerged for the factor language. In the foreign language condition (English), the assigned blame was *M=1.7416 (SD=0.8648),* and in the native language condition (Spanish), the blame assignment was *M=1.4882 (SD=0.7159),* with *F(1, 591)=14.77, p<.001.* The results contradict the hypothesis, as the native language condition assigned less blame to Damian on average than the foreign language condition.

Hypothesis 3

The interaction effect between the mutable condition and foreign language was not statistically significant*, F(1, 591)=2.98, p=.085.* The effect suggests that language did not have a significant effect on blame ratings between the mutable and immutable conditions. Although participants, on average, blamed Damian more in the mutable condition *(M=1.68, SD=0.89)* than in the immutable condition *(M=1.54, SD=0.70)*, it can not be concluded that the effect of mutability on blame was significantly different between the foreign and native language conditions.

Table 1

Descriptives of blame assignment, Damian

	Language					
	English (FL*)			Spanish (NL*)		
Condition	M	SD	N	M	SD	N
Mutable	1,87	0.99	155	1,50	0,73	159
Immutable	1,61	0,68	143	1,47	0,72	138
Total	1,75	0,87	298	1,49	0,72	297

*FL = Foreign Language, NL = Native Language

Hypothesis 4: **Halo effect**

Prior to the analysis, the assumptions were checked; homogeneity of variance was met with the Levene's test (p>.05), and the patterns of the histograms (Appendix B: Figures 4 and 6) and Q-Q plot (Appendix B: Figures 5 and 7) showed no major deviation from normality.

A 2 x 2 mixed ANOVA analysis was conducted to examine whether the trait ratings varied between the native and foreign languages. The within-subjects factor was the trait type (related vs. unrelated traits), and the between-subjects factor was the language condition

(Spanish vs. English). There was a significant main effect of trait type, $F(1, 593)=500.24$, $p<.001$, suggesting participants rated related traits *(M=3.94, SD=0.80)* more positively than unrelated traits *(M=3.16, SD=0.57).* Whereas there was no significant main effect of language, *F(1, 593)=0.94, p=.333,* meaning that overall trait ratings did not differ across the native *(M=3.57, SD=0.68)* and foreign *(M=3.53, SD=0.69)* language conditions. The interaction between trait type and language was significant, *F(1, 593)=8.26, p=.004,* indicating that participants in the foreign language condition rated unrelated traits *(M=3.09, SD=0.60*) significantly different from participants in the native language condition *(M=3.23, SD=0.54),* while related traits remained similar *(*foreign*: M=3.97, SD=0.80;* native*: M=3.91, SD=0.81)*.

Pairwise comparisons of the estimated marginal means were conducted, providing evidence that for unrelated traits, the differences between the foreign language condition *(M=3.09, SE=0.032)* were statistically significantly lower than in the native language condition *(M=3.24, SE=0.033), F(1, 610)=10.51, p=.001.* For related traits, no significant language difference was found *(English: M=3.97, SE=0.045; Spanish: M=3.91, SE=0.046), F(1, 610)=0.75, p=.386.* The difference between related and unrelated traits was significantly different in both language conditions: English *(F(1, 610)=336.24, p<.001),* and Spanish *(F(1, 610)=190.33, p<.001),* while the foreign language condition had a larger mean difference ($\Delta=0.88$) than the native language condition ($\Delta=0.68$).

Table 2

Descriptives of Trait ratings, Nora

	Language								
	English (FL*)			Spanish (NL*)			Total		
Trait type	M	SD	N	M	SD	N	M	SD	N
Related	3.97	0.8	298	3.92	0.81	297	3.94	0.80	595
Unrelated	3.08	0.59	298	3.23	0.54	297	3.16	0.57	595
Total	3.53	0.69	298	3.68	0.68	297			

*FL = Foreign Language, NL = Native Language

Discussion

This study investigated how blame attributions potentially differ in a foreign language and how that difference applies to counterfactual scenarios. Additionally, it was tested whether the foreign language effect could mitigate the halo effect.

When blame is assigned intuitively, it relies on heuristics of affective perception that lead to the disregarding of rational evaluation of blame, which could result in higher blame assignments (Skarlicki et al., 2017). As the use of a foreign language reduces heuristic and affective thinking and disconnects linguistic associations, we hypothesized that using a foreign language would lead to lower blame assignments (Costa et al., 2014). The results of our study do not support this hypothesis and indicate contrary effects; participants assigned higher levels of blame in their foreign language than in their native language.

Counterfactual thinking can increase the perception of perceived intention and the causal link between the actor and the outcome; therefore, we hypothesized that participants assigned to apply an if-only thought would give higher ratings of blame (Spellman & Gilbert, 2014). In the initial results (two-way ANOVA), participants in the mutable (vs immutable) condition applied significantly higher ratings of blame. Due to violations of the checked

assumptions, a second analysis was conducted (non-parametric test), which revealed non-significant results, meaning the hypothesis was not confirmed. To test whether the foreign language effect applies to counterfactual thoughts, we hypothesized that participants in a counterfactual scenario presented in their foreign language would assign lower levels of blame than participants in a counterfactual scenario presented in their native language. The analysis results in non-significant differences, therefore not supporting the hypothesis.

As the halo effect stems from heuristic thinking that overgeneralizes the perception of a central trait to other personality traits, processing in a foreign language could mitigate these processes (Keysar et al., 2012; Thorndike, 1920). In accordance with the hypothesis, based on describing "Nora" as "hard-working," participants rated unrelated traits in a foreign language (opposed to the native language) condition significantly lower. Presumably, through the reduced emotional affect and increase in analytical thinking, the bias of overgeneralizing had been mitigated. Although "Nora" was highly blamed for the negative outcome, it was not severe enough to diminish her overall positive perception and the resulting halo effect, as both language conditions exhibited significant overarching judgment on all trait evaluations.

Limitations, Strengths, and Future Directions

To accurately contextualize the findings of this study, it is essential to consider both its limitations and strengths. For the primary analysis of the Foreign Language effect in counterfactuals, the assumption of normality was mildly violated, which reduced the robustness of the initial finding that counterfactual thinking significantly increased blame scores. In the non-parametric analysis, these results were not replicated. Contributing to the distortion of normality could be the skewed distribution of blame scores, indicating a possible "floor effect," as the overall blame score for “Damian” in the scenario was relatively low (M=1.62), meaning that the majority of participants did not perceive Damian in the scenario as being at fault. Contrarily, we can observe a "ceiling effect" as participants blamed “Nora”

highly for her failing (M=4.35). Future studies should create a more ambiguous scenario in which the person to blame is less apparent to the participant.

Nevertheless, this study investigates the extensive effects of the foreign language effect on blame judgment in a counterfactual scenario and biased thinking, specifically on the Halo effect. Thereby, we provided additional insight into the underlying mechanisms of processing and their influence on human judgment. As the current study only assesses the halo effect in a foreign language after applying blame, future research could investigate how the halo effect varies depending on the severity of the blame.

The measurement of blame attribution was not supported by a validated scale. Although the two items assessing blame directly asked to evaluate blame, future research could enhance the construct validity of blame by using a validated questionnaire. The study employed an experimental design, enabling causal inferences about the effects measured between the foreign language condition and the native language condition. The simple random sample, with a large sample size (N=595), ensures relatively high statistical power and reliability for our findings. With the sample comprising Spanish-speaking participants, mainly from Mexico and Spain, the linguistic and cultural distances strengthen the internal validity of measuring the foreign language effect. As Spanish was the native language and English the foreign language, the results should be generalized with caution to other linguistic contexts. Future research should investigate the foreign language effect in other linguistic combinations to establish greater external validity.

Overall, this study could serve as a starting point for future research to address this study's limitations and utilize our findings to investigate how the foreign language effect influences the halo effect and blame assignments in counterfactual thoughts.

Theoretical and Practical Implications

The foreign language effect on blame

The foreign language effect builds on the premise of the reduced affect hypothesis, as termed by Keysar et al. (2012), which posits that emotional resonance is reduced in a foreign language. In our study, the participants reported a moderately low level of emotional affection towards the scenario, while there was no significant difference in emotionality between the foreign and native languages. In terms of the dual process theory (Barouillet, 2011), blame assignment tends to increase with heuristic thinking (System 1), which can be triggered by emotional arousal. The increased emotionality can lead to swift blame assignments, causing the rater to overestimate the intention behind the observed behavior and outcome, as well as the perceived causal control (Skarlicki et al., 2017). As this increased emotionality was not present in our study, it is likely that both conditions (foreign vs. native language) could equally apply rational attribution of blame (System 2). Thus, the foreign language effect could not lower blame by reducing emotionality and promoting analytical thinking (Alicke, 2000).

Contrary to the hypothesis, the participants in their foreign language were significantly more likely to attribute blame. An explanation for this effect could be that the reduced affect in a foreign language decreased empathy for "Damian," which led to less perspective-taking and, thereby, higher assignments of blame. Supported by recent research, empathy is found to be weaker when processed in a foreign language (Ward & Malgorzata Ragosko, 2025). This relationship must be further investigated before inferring the practical effect of the foreign language on attributing blame in clinical or social settings.

Counterfactuals and blame

Participants in both conditions reported high plausibility for the scenario, suggesting that those in the mutable condition were able to mentally simulate the counterfactual outcome

effectively. High plausibility aligns with the simulation heuristic, which proposes that the ease with which one can imagine and simulate an alternative outcome influences counterfactual reasoning and causal judgments, possibly increasing blame (Kahneman & Tversky, 1982). In the mutable condition, participants were more likely to perceive the outcome as changeable and identify mutable elements that can amplify the perception of causality, intention, and responsibility of an actor, possibly leading to an increase in blame assignment, especially when the alternative sequence alternative outcome appears easily attainable or emotionally salient (Alicke, 2000). Our data suggest a small effect, which should be interpreted with caution, as the additional analysis did not yield significant results. Blame was attributed quite differently between the two actors; for "Damian," participants generally disagreed with attributing blame (M=1.62), while for "Nora," the participants largely agreed to attribute blame (M=4.35). As blame primarily consists of identifying the responsible actor and assigning accountability, the consensus on blaming "Nora" suggests their perceived responsibility for the outcome was unambiguous. The influence of upward counterfactuals on blame also depends on the level of norm-breaking behavior described, as well as causal ambiguity, which tends to increase when the causal link is uncertain, allowing counterfactual reasoning to provide a new perspective (Spellman & Gilbert, 2014).

Counterfactual blame attribution in a foreign language

The expected interaction of thinking in a foreign language, which was predicted to cause decreased blame attributions in a counterfactual scenario, did not occur. Possibly due to flaws in the study's design, a lack of moral ambiguity, and low emotional involvement, the effects in the mutable and foreign language conditions were non-significant.

The halo effect in a foreign language

When participants were presented with a description of a central trait, a significant halo effect was observed in both language conditions, with a notable reduction in the halo

effect for unrelated traits in the foreign language condition compared to the native language condition. The heuristic thinking (System 1) that leads to intuitive and affective judgment, responsible for the halo effect, is mitigated in a foreign language by creating psychological and emotional distance. Enhanced analytical thinking (System 2) relies more on rational evaluations, which also weakens affective and contextual associations (Barouillet, 2011).

The traits related to the central trait "hard-working" (e.g., disciplined, ambitious) were rated as more likely to be present than unrelated traits (e.g., kind, charismatic), exhibiting a stronger halo effect. Related traits, amplified by semantic and logical associations, did not differ significantly across the language conditions (Westbury & King, 2024). In contrast, the halo effect for unrelated traits was significantly rated lower in a foreign language. Unrelated traits are primarily amplified through the affective general impression of a person, through which the rater judges and overgeneralizes traits. The foreign language effect reduces affective processing and disconnects affective linguistic associations that are thought to contribute to the overarching evaluation of unrelated character traits (Keysar et al., 2012).

The resulting diminished halo effect for people in their foreign language may also apply to organizational settings, such as hiring and performance review processes, by increasing rational judgment.

Conclusion

Overall, this study investigated the impact of the foreign language effect on judgments. While the results did not support the initial hypothesis that the foreign language effect would decrease blame, the contrary but significant results can be utilized for future research to investigate how the foreign language effect interacts with blame assignments, considering the role of empathy. A lack of emotional involvement and moral unambiguity due to the study’s design can, to some extent, explain the mixed results on the influence of counterfactual reasoning and increased blame assignments. Lastly, judgments in a foreign

language based on a single positive trait were significantly lower for unrelated traits, exhibiting a decreased halo effect. This supports the dual-process theory and reduced affect hypothesis that explains how the foreign language effect can mitigate cognitive biases (Keysar et al., 2012).

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574. https://doi.org/10.1037//0033-2909.126.4.556

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable Control and Counterfactual Reasoning in the Psychology of Blame. *Personality and Social Psychology Bulletin*, *34*(10), 1371–1381. https://doi.org/10.1177/0146167208321594

Alicke, M. D., & Zell, E. (2009). Social Attractiveness and Blame. *Journal of Applied Social Psychology*, *39*(9), 2089–2105. https://doi.org/10.1111/j.1559-1816.2009.00517.x

ARTEde. (2024, November 7). *Der letzte Sommer der DDR - Aufbruch und Anarchie | Doku HD | ARTE*. YouTube. https://www.youtube.com/watch?v=J-dgcJYuo7 I

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. *Developmental Review*, *31*(2-3), 79–85. https://doi.org/10.1016/j.dr.2011.07.002

Bodig, E., Toivo, W., & Scheepers, C. (2019). Investigating the foreign language effect as a mitigating influence on the “optimality bias” in moral judgements. *Journal of Cultural Cognitive Science*, *4*(2), 259–273. https://doi.org/10.1007/s41809-019-00050-4

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your Morals Depend on Language. *PLoS ONE*, *9*(4), e94842. https://doi.org/10.1371/journal.pone.0094842

Díaz-Lago, M., & Matute, H. (2018). Thinking in a Foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, *72*(1), 41–51. https://doi.org/10.1177/1747021818755326

Epstude, K., & Roese, N. J. (2008). The Functional Theory of Counterfactual Thinking. *Personality and Social Psychology Review*, *12*(2), 168–192. https://doi.org/10.1177/1088868308316091

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. *Judgment under Uncertainty*, 201–208. https://doi.org/10.1017/cbo9780511809477.015

Keysar, B., Hayakawa, S., & An, S. (2012). Psychological Science The Foreign-Language Effect : Thinking in a Foreign Tongue Reduces Decision Biases On behalf of: Association for Psychological Science. *Sage Journals*, *23*(6). https://doi.org/10.1177/0956797611432178

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, *46*(8), 819–834. https://doi.org/10.1037/0003-066x.46.8.819

Miozzo, M., Navarrete, E., Ongis, M., Mello, E., Girotto, V., & Peressotti, F. (2020). Foreign language effect in decision-making: How foreign is it? *Cognition*, *199*, 104245. https://doi.org/10.1016/j.cognition.2020.104245

Murray, S., O'Neill, K., Bridges, J., Sytsma, J., & Irving, Z. C. (2024). Blame for Hum(e)an Beings: The Role of Character Information in Judgments of Blame. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506241233708

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, *11*(2), 147–164. https://doi.org/10.1017/s1366728908003283

Roese, N. J., & Epstude, K. (2017). The Functional Theory of Counterfactual Thinking: New Evidence, New Challenges, New Insights. *Advances in Experimental Social Psychology*, *56*, 1–79. https://doi.org/10.1016/bs.aesp.2017.02.001

Skarlicki, D. P., Kay, A. A., Aquino, K., & Fushtey, D. (2017). Must Heads Roll? A Critique of and Alternative Approaches to Swift Blame. *Academy of Management Perspectives*, *31*(3), 222–238. https://doi.org/10.5465/amp.2015.0118

Spellman, B. A., & Gilbert, E. A. (2014). Blame, Cause, and Counterfactuals: The Inextricable Link. *Psychological Inquiry*, *25*(2), 245–250. https://doi.org/10.1080/1047840x.2014.901708

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29. https://doi.org/10.1037/h0071663

Ward, R., & Malgozata Ragosko. (2025). Does language experience and bilingualism shape empathy and emotional intelligence? *International Journal of Bilingualism*. https://doi.org/10.1177/13670069241308078

Weiner, B. (1995). *Judgments of responsibility: a foundation for a theory of social conduct*. Guilford Press.

Westbury, C., & King, D. (2024). A Constant Error, Revisited: A New Explanation of the Halo Effect. *Cognitive Science*, *48*(12). https://doi.org/10.1111/cogs.70022

Acknowledgments

I acknowledge the use of Grammarly (https://app.grammarly.com/) in helping me review my grammar and improve phrasing. As well as to Perplexity (https://www.perplexity.ai/) to help with the literature search, by using prompts as “Can you give me articles that discuss the role of heuristic thinking in the halo effect”. I used the output to read the suggested articles.

Appendix A

1. ***General Information about the Research:***

Why do I receive this information?

You are invited to participate in the research "Reflecting on what could have been.". This information is to inform you about what you can expect from this specific research. The topic of the research is how people evaluate unexpected situations in daily life and how they view the people involved. This research plan has been approved by the Ethical Committee Psychology. This research is conducted by Kai Epstude, Naomi Meijnen, Britt Hollander, Elias Eckardt, Monja Rinke, Anna de Wind, & Lara Bakker (University of Groningen)

Do I have to participate in this research?

Participation in the research is voluntary. However, your consent is needed. Therefore, please read this information carefully. If there is something unclear, or you have questions about the research, please contact k.epstude@rug.nl before participating. Only afterwards you decide if you want to participate. If you decide to not participate, you do not need to explain why, and there will be no negative consequences for you. You have this right at all times, including after you have consented to participate in the research.

Why this research?

The topic of the research is how people evaluate unexpected events in daily life and how they view the people involved. It will help to get a fuller understanding of how individuals form impressions of such situations.

What do we ask of you during the research?

After agreeing to the information and signing the informed consent, an online questionnaire will take place. There are no wrong answers, and you should not have to think about any negative consequences for any of your answers. Please answer the questions in the

questionnaire based on your own opinion. The research will take around 10 minutes of your time. You will be compensated via Prolific.

What are the consequences of participation?

As a participant in this study, you will be compensated via Prolific. There are no known risks to participation beyond those encountered in everyday life and the time that will be invested.

How will we treat your data?

Your data will be collected from the questionnaire. The measured traits cannot be linked to any personal information. We will have no direct access to directly identifying information (names, email addresses) and we will ensure privacy by keeping the data confidential. The only personal information that will be asked for during data collection are age, nationality and gender. We are not able to identify individual participants based on that, given that we have no access to information like names and emails for participants in the Prolific participant pool.

What else do you need to know?

You may always ask questions about the research: now, during the research, and after the end of the research. You can do so by emailing (k.epstude@rug.nl). Do you have questions or concerns regarding your rights as a research participant? For this you may also contact the Ethics Committee of the University of Groningen: ecp@rug.nl.

Do you have questions or concerns regarding your privacy, or regarding the handling of your personal data? For this you may also contact the Data Protection Officer of the University of Groningen: privacy@rug.nl.

As a research participant you have the right to a copy of this research information.

2. *Informed Consent (for Participants Aged 18 Years or Older)*

I have read the information about the research. I have had enough chances to ask questions about it. I understand that at the end of the research, there is room for further questions. I understand what the research is about, what is being asked of me, which consequences participation can have, how my data will be handled, and what my rights are. I understand that participation in the research is voluntary. I understand that my responses will be kept confidential. I myself choose to participate. I can stop participating at any moment. If I stop, I do not need to explain why. Stopping will have no negative consequences for me. By proceeding to the next page, I acknowledge that I have read the information and consent to participate in the study. (Please close this window if you do not want to participate in this research).

o Yes, I give my consent and wish to proceed.

3. Demographic Information

1. Which gender identity do you most identify with?

o Male

o Female

o Non-binary

o Prefer not to say

2. What is your current age?

3. What is your nationality?

o Spain

o Mexico

o Other _____

4. Which country do you currently reside in?

o Spain

o Mexico

o Other _____

5. Is the native language of your parents Spanish?

o Yes, of both my parents

o Yes, of my mother

o Yes, of my father

o No

o Other _____

6. Is the native language of your parents English?

o Yes, of both my parents

o Yes, of my mother

o Yes, of my father

o No

o Other _____

4. Language Characteristics

We would now like to ask you to answer some questions about using the English language.

1. How often do you use English in your daily life?

o Never

o Rarely

o Sometimes

o Often

o Daily

2. How would you describe your English proficiency?

o Terrible

o Poor

o Average

o Good

o Excellent

3. At around what age did you actively start to learn English? (In Years)

4. Where did you acquire most of your English skills? (Multiple answers possible)

- Formal education (school/university)
- Self-study (books, apps, online courses, etc.)
- Living in an English-speaking country
- Speaking with native speakers (friends, family, colleagues, etc.)
- Media (movies/shows, music, video games)
- Using English at work/daily life
- Other: _____

5. Instructions for Reading

Now, please read the scenario presented on the next page carefully.

Scenario Part: 1

Nora is a student who is studying for her final exam of her dream study. It is really important for her to get a good grade in order to graduate this year and get into a master's program. Nora is a hard-working student who is passionate about the course. She has been keeping up with the material the entire semester. For this last exam she needs some help with studying for this exam from her reliable tutor Damian. He has helped her in the past too with several courses.

Damian is her tutor. Throughout Nora's study he has helped her for several courses. A week before Nora's last exam he received an email from her asking for his help. He saw the email coming in, but did not open it because he had his mind somewhere else. This was because he was playing a game on his phone. Nora never received his help.

Scenario Part: 2

(Mutable Outcome Condition)

Nora has waited for two weeks and today she got her grade back from the exam. She clicks on the link online and finds out she failed the important course. Nora is devastated, because this means she will not be able to graduate this year. However, upon going to the exam review, she realized that she would have definitely passed with the help from Damian.

(Immutable Outcome Condition)

Nora has waited for two weeks and today she got her grade back from the exam. She clicks on the link online and finds out she failed the important course. Nora is devastated because she will not be able to graduate this year. However, upon going to the exam review, she realized that she would not have passed even with the help from Damian.

6. Timer

All of the following questions were timed: Timer 1 measured the first click, Timer 2 the last click, Timer 3 when the page was submitted and the click count was also measured.

7. Moral Judgement and Blame Assignment Questions

How much do you agree with the following statements? Click to the next page for each new statement.

These questions were measured on a Likert Scale from 1 to 5, with the following answer options: 1) strongly disagree, 2) somewhat disagree, 3) neither agree nor disagree, 4) somewhat disagree and 5) strongly agree:

1. Damian is to blame for Nora's bad grade
2. Nora is to blame for her bad grade
3. Nora could have expected Damian's help
4. Damian should have helped Nora
5. I feel bad for Nora

These questions were measured on a Likert Scale from 1 to 5, with the following answer options: 1) not at all, 2) slightly, 3) moderately, 4) very and 5) extremely:

6. Damian is at fault for Nora's bad situation
7. How emotional/personal did the scenario feel to you? *(Emotionality check)*

8. Measurement of Halo Effect

Based on the scenario describing Nora as a "hard working", how likely is it that she possesses the following character traits?

These questions were measured on a Likert Scale from 1 to 5, with the following answer options: 1) not at all, 2) slightly, 3) moderately, 4) very and 5) extremely:

1) Kind, 2) Dedicated, 3) Ambitious, 4) Adventurous, 5) Empathetic, 6) Disciplined, 7) Reliable and 8) Charismatic.

These questions also contained a control check: 'Pick "slightly likely" for this question.'

9. Measurements of Correspondence Bias and Likeability

How much do you think the following statements apply to the scenario?

These questions were measured on a Likert Scale from 1 to 5, with the following answer options: 1) not at all, 2) not really, 3) neutral, 4) somewhat and 5) very much.

The questions measuring either the correspondence bias or the likeability scores of Damian alternated and also included a control check and a plausibility check:

1. Damian's failure to help was due to his personality. *(Correspondence bias)*
2. I think Damian is a likeable person. *(Likeability Damian)*
3. Damian's failure to help was due to the situation. *(Correspondence bias)*
4. Damian's behavior reflects a general pattern of his personality. *(Correspondence bias)*
5. I believe Damian is a sympathetic person. *(Likeability Damian)*
6. External factors influenced Damian's decision. *(Correspondence bias)*
7. Pick "Somewhat" for this question *(Control check)*
8. I have warm feelings about Damian. *(Likeability Damian)*
9. I would blame myself for not responding, if I was in a situation like Damian. *(Hypothetical self-blame)*
10. The scenario was plausible/realistic to me. *(Plausibility check)*

10. Final Page

We thank you for your time spent taking this survey. Your response has been recorded.

Please click "continue" to return to Prolific.

Appendix B

Figure 1

Histogram of residuals (Blame)

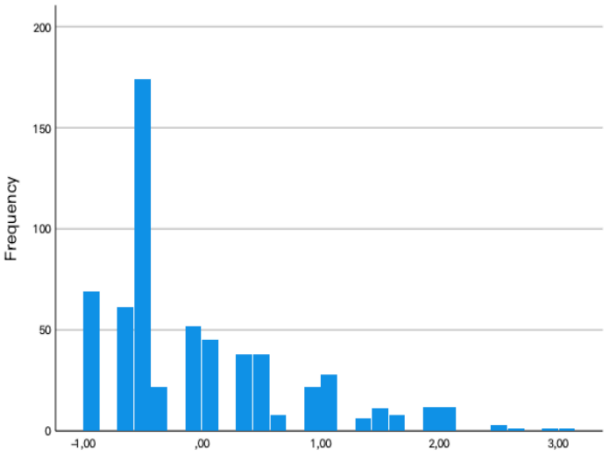

Figure 2

Q–Q plot of residuals (Blame)

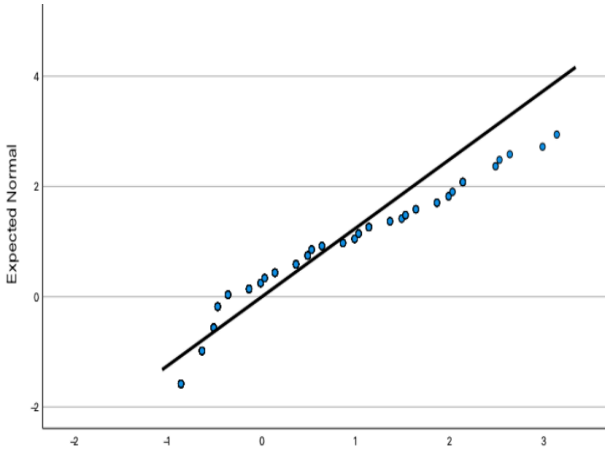

Figure 3

Boxplot of Blame assignment by Language × Condition

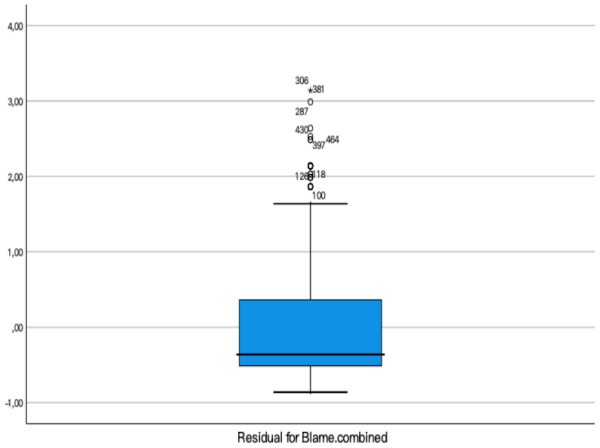

Figure 4

Histogram of residuals – Unrelated traits

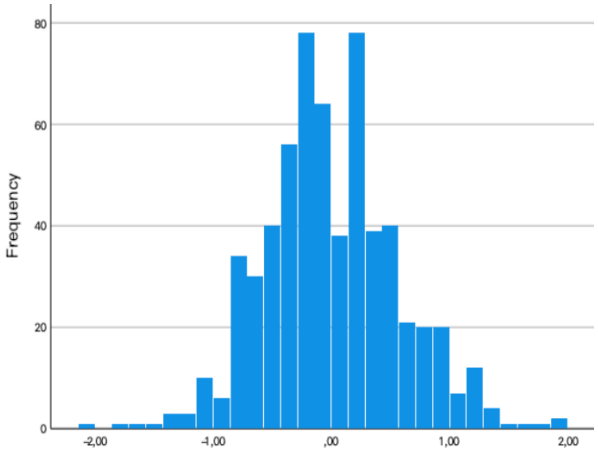

Figure 5

Q–Q plot of residuals – Unrelated traits

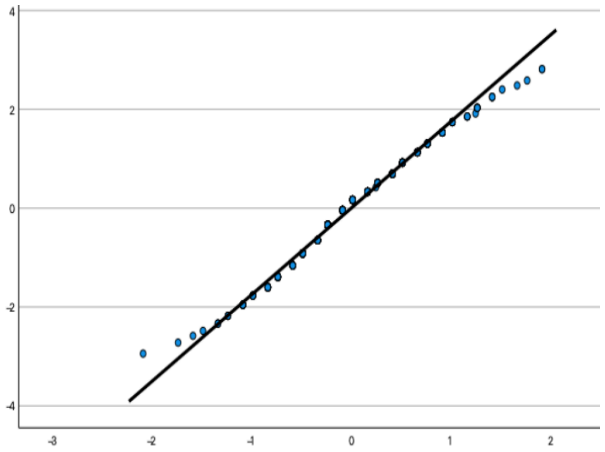

Figure 6

Histogram of residuals – Related traits

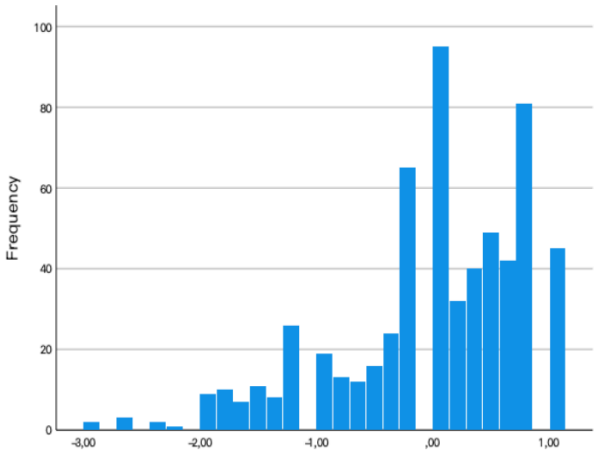

Figure 7

Q–Q plot of residuals – Related traits

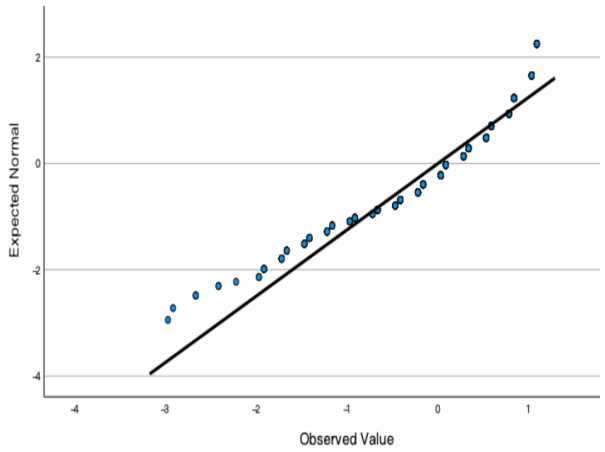

Table 3

ANOVA for blame assignments

Source	SS	Df	MS	F	Sig	Partial eta squared
Intercept	1540,56	1	1540,56	2468,35	<,001	0,807
Condition	3,08	1	3,08	4,93	,027	0,008
Language	9,22	1	9,22	14,77	<,001	0,024
Condition x Language	1,86	1	1,86	2,98	,085	0,005
Error	368,86	591	0,62			
Total	1935,50	595				

Table 4

Additional Descriptives

		Language								
		English (FL*)			Spanish (NL*)			Total		
Variable	Condition	M	SD	N	M	SD	N	M	SD	N
Blame Nora	Mutable	4.35	0.82	162	4.40	0.85	162	4.37	0.84	324
	Immutable	4.24	0.85	145	4.41	0.67	138	4.33	0.77	283
	Total	4.30	0.84	307	4.40	0.77	300	4.35	0.81	607
Sympathy for Nora	Mutable	3.88	1.12	162	3.84	1.15	162	3.86	1.13	324
	Immutable	4.25	0.86	145	3.91	1.06	138	4.08	0.98	283
	Total	4.05	1.02	307	3.87	1.11	300	3.96	1.07	607
Emotionally affected	Mutable	2.75	1.02	162	2.81	0.95	162	2.78	0.98	324
	Immutable	2.85	0.95	145	3.01	1.01	138	2.93	0.98	283
	Total	2.80	0.99	307	2.90	0.98	300	2.85	0.98	607
Plausibility of the scenario	Mutable	4.17	0.76	162	4.68	1.23	162	4.42	1.05	324
	Immutable	4.16	0.81	145	4.76	1.14	138	4.45	1.03	283
	Total	4.16	0.78	307	4.72	1.19	300	4.44	1.04	607

*FL = Foreign Language, NL = Native Language