Title Page (containing all authors contact details)

Click here to access/download;Title Page (containing all authors contact details);20231122_Bribery_FLE_Title_Page.docx ⬇

Is There a Foreign Language Effect on Workplace Bribery Susceptibility? Evidence from a Randomized Controlled Vignette Experiment

Running Head: Is There a Foreign Language Effect on Workplace Bribery Susceptibility?

Jack Fitzgerald*[1], Paul Stroet[†2], Kristina S. Weißmüller[†3], Arjen van Witteloostuijn[†1, 4]

[1] Department of Ethics, Governance, and Society, School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1085HV Amsterdam, the Netherlands

[2] Department of Management and Organization, School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1085HV Amsterdam, the Netherlands

[3] Department of Political Science and Public Administration, Faculty of Social Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1085HV Amsterdam, the Netherlands

[4] Faculty of Applied Economics, Antwerp Management School, University of Antwerp, Boogkeers 5, 2000 Antwerp, Belgium

*Corresponding author. Email: j.f.fitzgerald@vu.nl. Phone: +31 (0)6 21 88 22 04.

[†]Contributing authors. These authors contributed equally to this work. Emails: p.stroet@vu.nl, k.s.weissmueller@vu.nl, a.van.witteloostuijn@vu.nl.

Declarations

Funding
No funding to declare.

Competing Interests
All authors declare that they have no competing interests with regard to this article and give their informed consent to its submission to the Journal of Business Ethics.

Ethics Approval
The Ethical Review Board of the School of Business and Economics at Vrije Universiteit Amsterdam granted ethical approval for this study. This research is performed in accordance with the Netherlands Code of Conduct for Research Integrity, a comparable ethical standard to the 1964 Declaration of Helsinki that applies to all scientific disciplines.

Informed Consent
We obtained informed consent from all respondents in the study prior to such respondents' participation.

Materials Availability
Code necessary to reproduce the tables, figures, and other empirical findings is available at https://doi.org/10.17605/OSF.IO/Y3NQ7. Data is available by request of the authors.

Authors' Contributions
Conceptualization: Jack Fitzgerald, Kristina S. Weißmüller, Arjen van Witteloostuijn; Methodology: Jack Fitzgerald, Kristina S. Weißmüller, Arjen van Witteloostuijn; Formal analysis and investigation: Jack Fitzgerald, Paul Stroet; Writing - original draft preparation: Jack Fitzgerald; Writing - review and editing: Kristina S. Weißmüller, Arjen van Witteloostuijn; Supervision: Kristina S. Weißmüller, Arjen van Witteloostuijn.

Is There a Foreign Language Effect on Workplace Bribery Susceptibility?

Evidence from a Randomized Controlled Vignette Experiment

Abstract

Theory and evidence from the behavioral science literature suggest that the widespread and rising use of *lingua francas* in the workplace may impact the ethical decision-making of individuals who must use foreign languages at work. We test the impact of foreign language usage on individuals' susceptibility to bribery in workplace settings using a vignette-based randomized controlled trial in a Dutch student sample. Results suggest that there is not even a small foreign language effect on workplace bribery susceptibility. We combine traditional null hypothesis significance testing with equivalence testing methods novel to the business ethics literature that can provide statistically significant evidence of bounded or null relationships between variables. These tests suggest that the foreign language effect on workplace bribery susceptibility is bounded below even small effect sizes. *Post hoc* analyses provide evidence suggesting fruitful further routes of experimental research into bribery.

Introduction

Many people around the world leave home speaking one language and arrive at work speaking another. Within the past two decades, already about 25% of European Union citizens used a foreign language (FL) at work (Eurobarometer 2012), and 35% of European small and medium-sized enterprises provided FL training to their employees (Hagen et al. 2006). English is a particularly dominant *lingua franca* in international business settings including travel, business associations, news, entertainment, technology, and publishing (Kankaanranta & Lu 2013; Melitz 2016). While language standardization is historically prevalent in multinational corporations (Marschan-Piekkari et al. 1999), *lingua franca* usage in the workplace is rising in businesses of all sizes due to globalization. The increasing demand for workers with FL competency across many regions in the world is strong evidence of this shift. For instance, 52% of job postings in the Visegrad region of Eastern Europe require English language proficiency (Fabo et al. 2017), and FL knowledge and proficiency respectively yield average wage premiums of 13% and 11% in the German labor market (Hahm & Gazzola 2022).

The behavioral science literature suggests that this widespread and rising use of *lingua francas* in the workplace may impact the ethical decision-making of individuals who use FLs at work. Prior studies find a behavioral phenomenon known as the **foreign language effect** (FLE) whereby people make systematically different decisions when evaluating choice problems in a native language (NL), compared to when those same problems are evaluated under FL usage (Keysar et al. 2012; Costa et al. 2014; Geipel et al. 2015; Hayakawa et al. 2017; Circi et al. 2021). The theoretical impetus for this effect is the challenge of FL usage, which activates more deliberative brain processing and crowds out simple heuristics that systematically bias decision-making (Keysar et al. 2012). Moral norms are one such group of heuristics that may become crowded out during FL usage.

This results in a commonly found **moral FLE**, whereby people become less deontological (and more utilitarian by comparison) when evaluating moral dilemmas under FL usage rather than NL usage (Costa et al. 2014; Geipel et al. 2015; Hayakawa et al. 2017; Circi et al. 2021; Stankovic et al. 2022). Such an effect is of obvious interest for understanding the determinants of (un)ethical behavior in the workplace. A potential link between *lingua franca* usage and unethical behavior such as corruption is particularly important to investigate, as corruption in international business is associated with a host of negative impacts (Bahoo et al. 2020). Chief among these impacts include worse firm performance (Lee & Hong 2012) and lower corporate responsibility (Keig et al. 2015).

The current study experimentally investigates whether such an FLE arises for arguably the most visible and well-studied form of corrupt behavior in workplace settings: bribery (Robertson & Nichols 2017). We use the composite Likert scale from De Waele et al. (2021) and Weißmüller & De Waele (2022) to build a measure of **workplace bribery susceptibility (WBS)**, which captures individuals' willingness to accept a bribe. We validate this measure for the bribery acceptance setting through exploratory and confirmatory factor analyses, and detail theoretical underpinnings for each component of the measure. We then employ a vignette-based randomized controlled trial (RCT), inspired by the same pair of papers, with a sample of business administration students at a large international Dutch university to assess how the language in which the scenario is processed (either Dutch, their NL, or English, their FL) impacts WBS. We opted for a vignette-based RCT because the use of realistic vignettes increases external validity, whilst the RCT design facilities credible causal inference.

The results suggest that there is not even a small FLE on WBS for our sample. Our analysis combines traditional null hypothesis significance testing (NHST) with equivalence testing methods that can provide statistically significant evidence that relationships between variables are bounded beneath given effect sizes. To our knowledge, this study is the first in

the business ethics literature to employ such equivalence testing. Our empirical evidence suggests that any FLE on WBS is bounded below even small effect sizes. Furthermore, results are inconclusive with regards to the question of whether FLEs vary for participants with different levels of FL proficiency. Both findings are striking in the face of the extant literature, which persistently finds significant moral FLEs that are moderated by FL proficiency (Circi et al. 2021; Stankovic et al. 2022). Further, *post hoc* analyses provide tentative evidence of several bribery hypotheses worth exploring in future studies. First, people may be significantly more susceptible to accepting than to offering bribes, and may be more susceptible to private-to-private bribery, which only involves private sector actors, than to public bribery, where at least one party in the transaction is a public sector actor. Second, individuals may be more willing to accept bribes that they perceive as being less harmful to the goals of their organization. These conclusions require further research for confirmation and understanding.

These novel insights offer four contributions to the literature. First, we add an FLE study to the tiny stream of work on this issue in business ethics (Pan & Patel 2018; Sugahara et al. 2023). FLE studies are increasingly needed in times of rising *lingua franca* usage across the world. This applies to business ethics particularly, given extant evidence regarding the moral FLE. Second, we provide a direct assessment of the FLE's replicability and generalizability in more realistic settings. FLEs are often tested in highly artificial dilemma vignettes which are highly unlikely to be experienced in real life. We show empirically that participants view our vignettes to be realistic on average, and thus our findings provide novel evidence on the FLE's existence (or lack thereof) in more realistic settings. Third, we open a conversation concerning the relevance of bribery susceptibility in contexts that only involve private sector actors. As we will further illustrate, specific types of unethical behavior involving only private sector actors can still be seen as important instances of bribery, given their harmful effects in the public domain. Fourth and finally, we add equivalence testing methods to the

4

business ethics research community's toolkit. These methods offer an extension to the well-known NHST approach and permit more credible publication of null effects, as equivalence testing methods can provide evidence of such null effects with some certainty of Type I error coverage. This is an important step forward to ensuring replicability and falsifiability of results in the business ethics literature, as recent literature on the replicability crisis in the social sciences (see, for example, van Witteloostuijn 2016; Hensel 2021; Youyou et al. 2023) makes clear that the research community must be much more open to publishing null findings.

This study is structured as follows. In the next section, we discuss the WBS measure. We also detail why bribery acceptance, as well as bribery that occurs entirely between private sector actors, are important research domains. Next, we detail the multidisciplinary literature behind the FLE, offer a cogent theoretical mechanism for the moral FLE, and use this theoretical background to build a pair of FLE hypotheses. Subsequently, we introduce the methodology and empirical design, discussing the sample, vignettes, and measures. After presenting the empirical results, the study concludes with a discussion of its theoretical contributions, limitations, and directions for further research.

Theory

Workplace Bribery Susceptibility

Ramdani & van Witteloostuijn (2014) note that classical academic definitions of bribery are quite restrictive, as they require the involvement of a public sector actor. Thus, though bribery is an extensively studied phenomenon (Robertson & Nichols 2017), bribery research tends to ignore bribery settings where no public sector actor is involved. Ramdani & van Witteloostuijn (2014) address this issue by providing a more expansive definition of bribery, including its business-to-business manifestation: "the corrupt payment, receipt, or solicitation of a private favor for actions or decisions from influential or powerful agents or

5

authorities which could be public officials, corporations or people inside corporations to generate private benefits for the briber." We adopt this broader definition for the purposes of this study.

We build off this definition, and uniquely add to extant bribery work, by examining bribery which occurs exclusively between private sector actors. There are many cases where such bribery is of considerable social interest. For example, Herics et al. (2018) document 1,749 public-private partnerships worth 336 billion euros in European Union Member States since the 1990s. Private contractors carrying out public functions in such partnerships who accept bribes for undue favors may waste considerable amounts of public funds or otherwise harm social welfare through corruption, and may do so without any direct involvement of a public official. Further, private sector individuals with significant decision-making power in their organizations who accept bribes in exchange for undue favors may act against the interests of their organization and/or against the public interest. Such corruption may constitute a violation of fiduciary responsibility (for instance, to shareholders and/or stakeholders), or outright fraud if the corruption is egregious enough. Additionally, witnesses may be deterred from whistleblowing on corporate crime via hush money payments. This suppression of incriminating information about company behavior serves as a considerable barrier to corporate accountability. These are all scenarios in which there is strong social and public interest in the descriptives and determinants of bribery susceptibility, and in which no public sector actor is involved. We thus use each of these cases as bribery scenarios in our experiment. This use of multiple bribery vignettes assists with the generalizability of our findings by ensuring that the FLE results found here are not simply an artefact of the chosen vignette scenario (Robertson & Nichols 2017).

The generalizability of findings relating to determinants of bribery susceptibility may be threatened by only focusing on bribery in the public sector for several reasons. Principally,

6

both bribery levels and treatment effects are expected to differ between public and private-to-private bribery. This is because both descriptive and prescriptive norms differ between the two forms of bribery. Private-to-private bribery is perceived to be as common, if not more common, than public bribery (Jaakson et al. 2019). Descriptive norms against private-to-private bribery are thus potentially weaker than those against public bribery. The same is certainly true of prescriptive norms against private-to-private bribery. Despite the large economic inefficiencies associated with private-to-private bribery (Argandoña 2003; Gopinath 2008; Goel et al. 2015), such bribery is still widely considered to be a 'victimless crime' (Jaakson et al. 2019). As a result, compared to public bribery, it is much less clear to potential offenders that private-to-private bribery is an ethical or legal violation (Gopinath 2008). These more lax prescriptive norms against private-to-private bribery even spill over into more lax legal enforcement (Argandoña 2003). For example, American anti-corruption legislation historically offers specific exemptions for 'facilitating payments', a form of private-to-private bribery used to facilitate the completion of private-sector transactions, in foreign countries (Argandoña 2005; Gopinath 2008). Given the clear negative relationship between anti-bribery norms and bribery susceptibility (for example, see Köbis et al. 2015; Banerjee 2016; Abbink et al. 2018; Senci et al. 2019), people should be expected to be more susceptible to private-to-private bribery than to public bribery, implying higher levels of private-to-private bribery than public bribery. Further, when treatment effects depend on individuals' heuristic adherence to norms – which is true of the FLE – the weaker norms against private-to-private bribery imply lower heuristic adherence to such norms, which may impact the effects of interventions designed to change such heuristic adherence.

Additionally, because private-to-private bribery attracts less scrutiny than public bribery due to an absence of public funds, private-to-private bribery is less visible than public bribery, and thus much less is known about private-to-private bribery than public bribery

7

(Argandoña 2003; Goel et al. 2015). Therefore, though the determinants of public bribery are well-studied (Robertson & Nichols 2017; Weißmüller & Zuber 2023), private-to-private bribery may exhibit vastly different determinants, and these determinants may not be well-captured in the few visible cases of private-to-private bribery, which may themselves be unrepresentative. Developing new data on private-to-private bribery, even in an experimental context, is thus critical to protecting against this threat to the external validity of bribery research.

Our central dependent variable is workplace bribery susceptibility (WBS), which measures an individual's propensity to accept a bribe. Pre-empting the methods section, we briefly discuss our measure to clearly and precisely introduce our definition of WBS. We construct our WBS measure from participants' responses to each of our vignettes, specifically using the same four-item Likert scale used to measure **willingness to bribe (WTB)** in De Waele et al. (2021) and Weißmüller & De Waele (2022). However, in contrast to these prior measures, our Likert items elicit participants' beliefs concerning the **acceptance** of the bribes in our vignettes, rather than the **offering** of such bribes (more details are provided in the methods section). This is an important shift in focus, as examining demand-side bribery determinants is just as important as examining supply-side bribery determinants in developing a comprehensive understanding of why bribery materializes.

Bribery acceptance is important to analyze not only because it is one of the two necessary sides of bribery, but also because there are strong theoretical reasons to believe that bribery levels and treatment effects differ between bribery acceptance and bribery offering. When one **offers** a bribe, they often do not know *a priori* whether the individual on the receiving end of this bribe will accept, reject, or even report the bribe. The descriptive norms in this setting are thus unclear, and a person offering a bribe assumes a considerable risk of punitive consequences for bribery. Even if the bribe is accepted and not reported, a person

8

offering a bribe still takes on a pecuniary cost; they must actually pay the bribe. However, when one **accepts** a bribe, reporting risks from the other party in the bribe are nearly zero, with the only significant risk arising from the potential of sting operations. Further, though bribe offering is costly, bribe acceptance confers direct pecuniary benefits. Additionally, facing a bribe offer can cause a potential bribe accepter to sharply update their beliefs on descriptive norms, as they are directly confronted by at least one person in their immediate social environment who believes that bribery is acceptable. Finally, because a bribe accepter is more passive than one who offers a bribe, they may engage in a form of ethical free riding (see Gross et al. 2018), resolving the cognitive dissonance that arises from engaging in corruption while maintaining a positive self-image by briefly adopting the ethical standards of a less ethical partner.

These factors jointly imply that people are more likely to accept a bribe than to offer one. This is true under both a rational choice framework and under a more behavioral, norm-centric framework. Under a rational choice framework, bribery acceptance confers lower risk and higher rewards than bribery offering, which implies that rational actors are more likely to accept a bribe than to offer one. Under a norm-centric framework, because descriptive norms against bribery are perceived to be lower when accepting a bribe than when offering a bribe, the aforementioned negative relationship between descriptive norms against bribery and bribery activity (see Köbis et al. 2015; Banerjee 2016; Abbink et al. 2018; Senci et al. 2019) also implies that people are more likely to accept than to offer bribes. Additionally, as is discussed in the next section, the lower risk and weaker norms associated with bribery acceptance compared to bribery offering may alter treatment effects that depend on heuristic moral or risk-related decision-making. As is discussed in the next subsection, this directly concerns the FLE.

9

The Foreign Language Effect

Many recent studies detect FLEs in decisions involving risk or ethical dilemmas. Meta-analytic evidence from Circi et al. (2021) shows that two FLEs are persistently found in the literature. The first is the **moral FLE** whereby people respond in a more utilitarian and less deontological fashion when evaluating decision problems under FL usage. That is, decision-makers appear to be more likely to prioritize collective well-being over individual ethical obligations under FL usage. For example, prior research finds that participants in experiments report being more willing to kill one person to save five under FL usage (Costa et al. 2014; Geipel et al. 2015; Hayakawa et al. 2017). The second FLE often found is the **risk FLE** whereby FL usage induces lower levels of risk aversion. While the risk FLE is certainly relevant to any potential FLE on workplace bribery decisions, we focus our theoretical discussion on the moral FLE because this is arguably the most consistent and well-documented effect examined in the FLE literature (Hayakawa et al. 2019; Stankovic et al. 2022). For example, more than 70% of the FLE experiments identified for the FLE meta-analysis in Circi et al. (2021) are moral FLE experiments.

While there exist multiple explanations for the FLE,[1] the most relevant and familiar mechanism is based on dual process theory – that is, the System 1/System 2 model of cognitive processing (Stanovich & West 2000; Kahneman 2011). Dual process theory postulates that humans approach most decision problems with fast, intuitive, and heuristic System 1 processing by default, and switch to slower, more deliberative, and more effortful System 2 processing only when faced with stimuli or problems which appear to require System 2 processing due to higher difficulty and/or complexity. The dual process theory explanation posits that FL usage increases cognitive load and makes individuals more likely to utilize System 2 processing in place of System 1 processing.

Dual processing may yield moral FLEs through two routes. The first potential route is through FL-induced improvements in cognitive reasoning (see, for instance, Keysar et al. 2012). However, recent findings show that this route is unlikely; FL usage does not increase people's capability of solving logical problems devoid of emotional context (Costa et al. 2014, Mækelæ & Pfhul 2019), and some cognitive biases actually increase under FL usage (for example, see van Hugten & van Witteloostuijn 2018). The second theoretical route is through FL-induced reductions in norm-based reasoning. If FL usage suppresses System 1 processing, then decision-making heuristics such as ethical norms may be crowded out. This theory is supported by recent research findings (Greene 2014; Geipel et al. 2015; Białek et al. 2019) and is much more plausible than a cognitive reasoning mechanism. The norm-based reasoning reduction mechanism explains why it is often found that common norms such as "thou shalt not kill" are less likely to be followed when participants consider killing one person to save five (Costa et al. 2014; Geipel et al. 2015), and also explains findings that while FL usage decreases deontological reasoning during such moral dilemmas, levels of utilitarian reasoning remain unchanged (Hayakawa et al. 2017).

While a utilitarian would then laud FL usage for 'improving' ethical decision-making (note that a deontologist would not reach the same conclusion), we argue that this observed effect arises only because of how ethical dilemma experiments are designed. Choice problems such as 'kill one to save five' posit two choices where reasonable people could disagree over which one is more ethical. This is not true of most ethical dilemmas faced in everyday life. The primary conflict in ethical decision-making is not usually between two schools of philosophy, but between self-interest and morality. Our workplace bribery setting is one such case where ethical choices are pitted against self-serving choices. Choice problems like these imply that the suppression of ethical norms would yield fewer ethical choices, and therefore more self-serving choices. We thus predict that FL usage will increase WBS by suppressing ethical norms

11

concerning bribery. The same hypothesis could also be reached through the mechanism of lower risk aversion under FL usage.

Hypothesis 1 (H1). *FL usage increases WBS.*

Given this mechanism, the theory and empirical evidence on the moral FLE also offer reasons to believe that the moral FLE differs by **FL proficiency (FLP)**, which is a standard control or moderator variable in prior FLE research (see, for example, Urbig et al. 2016; Stankovic et al. 2022). This is because higher FLP yields lower cognitive load from FL usage. The dual process theory account of the moral FLE thus suggests that individuals with higher FLP experience a smaller moral FLE because such participants are less likely to activate System 2 processing when switching to the FL, and are therefore more likely to heuristically rely on ethical norms. Indeed, meta-analytic results demonstrate that higher proficiency in the treated FL negatively moderates the observed moral FLE (Stankovic et al. 2022). We hypothesize that the same effect will be observed in our data.

Hypothesis 2 (H2). *The FLE on WBS will be negatively moderated by FLP. Specifically, the FLE on WBS will attenuate toward zero for participants with higher levels of FLP.*

Moral FLEs have received very modest attention in the business ethics literature (see, for instance, McDonald 2000; Warner et al. 2022). Specifically, experimental moral FLE research in business ethics to date is (to our knowledge) restricted to analyzing how ethical accounting judgments vary under FL usage (Pan & Patel 2018; Sugahara et al. 2023). The present moral FLE literature in business ethics is thus quite limited. The key concern of these accounting studies is the fact that the English-language International Financial Reporting Standards (IFRS) have become the *de facto* global *lingua franca* for accounting standards. These studies are thus interested in whether the ethical standards therein are interpreted and applied differently under FL usage. The moral FLEs we analyze – those of workplace bribery

12

scenarios – are less vulnerable to confounding[2] and much more generalizable to other forms of

real-world ethical decision-making.

Material and Methods

Our experimental protocol differs from most in the FLE literature. By far the majority of FLE experiments treat participants by changing the language of the participant's survey, including portions of the survey where pre-treatment covariates are collected (see, for example, Costa et al. 2014; Hayakawa et al. 2019). However, the FLE is widely postulated to impact numerous cognitive processes and decision-making tasks (Circi et al. 2021, Stankovic et al. 2022). Any 'pre-treatment' covariate that the FLE influences is in fact collected under treatment when collected using a survey with a randomly-assigned language. This opens an often-ignored risk that 'pre-treatment' covariates are actually 'bad controls', and hence bias the treatment effect(s) of interest when incorporated into statistical models (Montgomery et al. 2018; Cinelli et al. 2022).[3]

We address this common design imperfection in the FLE literature by conducting our experiment in two parts. The first part is a pretreatment survey (conducted in the NL) where we collect pre-treatment covariates. We choose the NL for this first part of the experiment to ensure that we are measuring confounders in the absence of treatment. The second part of the experiment introduces our treatments in a 3x2 factorial design. Participants randomly receive the second survey in either Dutch (NL) or English (FL) and are randomly presented with two of three bribery vignettes: the *inspector* vignette, the *negotiator* vignette, and the *whistleblower* vignette. These vignettes are detailed in the next subsection. We protect against experimenter demand biases by administering the second part of the experiment to participants one week after the first part is completed. All treatments are randomly assigned with equal probability. Presenting multiple vignette treatments to each participant improves power but implies that responses are clustered at the individual level. We thus estimate cluster-robust standard errors across all regressions to accommodate this clustering.

14

Vignettes

The inspector, negotiator, and whistleblower vignettes are selected to emulate real scenarios where bribery between private sector actors is societally relevant. The inspector vignette concerns an employee of a private firm who performs health inspections in restaurants. The inspector is offered a free dinner for themselves and their family at their favorite restaurant if they waive that restaurant's regular inspection. The negotiator vignette concerns a private-sector employee tasked with traveling abroad to negotiate a new contract with a foreign supplier. The employee finds the keys to a sports car (which the employee can use for the duration of their stay) on their hotel bed with a note stating that this gift is a token of good will in anticipation of more cooperation between the supplier and the employee's company. The whistleblower vignette concerns a whistleblower at an international shipping company tasked with testifying against their department's supervisor for taking bribes in exchange for tip-offs to drug cartels when package inspections were to take place. The whistleblower receives a letter offering 10,000 euros in exchange for not testifying.

All three vignettes reflect plausible real-world scenarios of bribery in the private sector that potentially harm the public interest. These vignettes are designed such that participants' self-reported likelihood of, comfort with, and beliefs about accepting the bribe in the vignette are indicative of their general susceptibility to accepting bribes at work. The full English-language vignette texts can be found in Appendix B. We can verify the accuracy of translations without external assistance as the research team includes native Dutch speakers who are proficient in English and a native English speaker who is proficient in Dutch. Notice that these vignettes vary greatly in the financial value of the bribe offer and the social consequences of accepting the bribe offer. We introduce this large variation intentionally, as this helps ensure that our moral FLE findings cannot be explained exclusively by the monetary value or social

consequences of accepting the bribe (given that our FLE findings are consistent across vignettes).[4]

Such bribery stakes are an important factor for analyses of bribery behavior. Qualitative and theoretical corruption research has long distinguished between white, grey, and black corruption (for example, see Heidenheimer 1970; Gardiner 2001; Werner 2001), and this classification has also been extended to bribery (Ramdani & van Witteloostuijn 2012). These 'shades' of bribery describe the severity of a particular bribery offense, with 'white' bribes being deemed the most acceptable bribes and 'black' bribes being deemed the least acceptable by a given society. The stakes and severity of a bribe, both for the direct participants in the bribe and for society at large, are key determinants of the bribe's shade. There is empirical evidence that bribery severity directly influences bribery predispositions: Weißmüller & De Waele (2022) and De Waele et al. (2021) both show that individuals' WTB is significantly and negatively monotonically related to darker bribery shades.

Bribery severity is expected to not only impact **levels** of bribery predisposition, but also **treatment effects** on bribery predisposition, including FLEs. The primary mechanism behind the FLE is a mental switch from heuristic System 1 thinking to deliberative System 2 thinking (see Keysar et al. 2012; Greene 2014; Geipel et al. 2015; Białek et al. 2019), so any confounder that predisposes individuals to one of those two states could bias FLEs. Bribery stakes are thus important for FLE estimation, as more extreme bribery states may trigger heuristic thinking. Sufficiently low bribery stakes can trigger 'no-big-deal' heuristic thinking whereby a bribery participant rationalizes their actions by considering the stakes to be so low that the bribe itself does not matter. Similarly, sufficiently high bribery stakes may signal to bribery participants that the bribe is socially unacceptable, and thereby may invoke social norms against black bribery, leading people to not engage in bribery due to heuristic adherence to such norms. It is

thus useful for external validity to ensure that our FLE findings hold after controlling for these different levels of bribery stakes.

Sample

Participants are recruited from the behavioral research lab of a large Dutch university. The recruitment pool consists of first- and second-year students from business administration bachelor programs. Study participation is voluntary and is incentivized with credit points for participants' study programs. Our surveys took place between November and December 2022.[5] The final sample is restricted to participants who report speaking Dutch as their first language, ensuring that our English language treatment is a reliable proxy for an FL treatment. Dutch first-language speakers constitute 89% of our sample. We ensure data quality by running attention checks. Participants are informed that they will not receive research participation credits for the experiment if they do not answer the attention checks correctly. The flow of study participants is visualized in Figure 1, which shows that observation attrition due to attention check failures amounts to a rather low 6.5% of the remaining sample.[6]

[Insert Figure 1 about here]

While data on self-assessed English language proficiency are collected, participants are not screened for a sufficient level of English proficiency as an inclusion criterion. This is because functionally all students in the recruitment pool either participate in an English-language bachelor program or hold a diploma from a Dutch secondary school, which implies high English language proficiency. English is the only compulsory FL in Dutch secondary schools since 1968 (Edwards 2016). The required level of English proficiency to graduate from high schools in the Netherlands is roughly B2-C1 level per the Common European Framework of Reference for Languages (College voor Toetsen en Examens 2023), and 90% of Dutch citizens across all educational levels are conversationally proficient in English (Eurobarometer

17

2012). Note that a potential disadvantage of our sample is low variation in FL proficiency. We return to this drawback as we discuss our results.

Internal Validity of the WBS Measure

We develop a valid and reliable measurement of our dependent variable, which is novel to the literature: WBS. Our WBS measure largely amounts to replacing references to "offering" in De Waele et al.'s (2021) and Weißmüller & De Waele's (2022) original bribery propensity Likert items with "accepting" in ours. The four items in our Likert measure represent the relevant dimensions of bribery susceptibility, specifically a participant's **likelihood** of accepting the bribe, **comfort** with accepting the bribe, degree of perceived **justification** for accepting the bribe, and the extent to which the participant believes that accepting the bribe would be a **mistake** (reverse-scored). As in De Waele et al. (2021) and Weißmüller & De Waele (2022), we also ask three questions as manipulation checks regarding the realism of the vignette, the perceived benefit of accepting the bribe to the briber within the vignette, and the perceived harm of the bribe to the participant's business organization in the vignette. We provide the exact English-language Likert items used to elicit our WBS measure components and our manipulation checks in Appendix A. Each of these items are asked as five-point Likert scales, as in the scales employed by De Waele et al. (2021) and Weißmüller & De Waele (2022).

We expand on the contributions of Weißmüller & De Waele (2022) and De Waele et al. (2021), and clarify the theoretical value of this scale for measuring bribery susceptibility, by detailing behavioral theoretical underpinnings for each of WBS' four scale components. First, eliciting the perceived likelihood of accepting a bribe effectively elicits intentions of unethical activity. Psychological theories of planned behavior posit that these intentions are a necessary antecedent to unethical behavior, and there is indeed empirical evidence of this connection (Huang & Chen 2023).

18

Second, eliciting participants' comfort with accepting bribes is a direct measure of their moral dissonance with this action (see Lowell 2011). This is an important predictor of unethical behavior, as cognitivist theories of criminology posit that people's desire to maintain their positive self-image and avoid moral dissonance is a key deterrent to criminal behavior (see Paternoster & Simpson 1996; Paternoster et al. 2015; Jaakson et al. 2019).

Third, eliciting participants' perceived justification of accepting bribes is a direct measure of their ability to neutralize this dissonance (see Sykes & Matza 1957). Such neutralization increases one's tolerance of unethical behavior, and can do so even when one's personal moral code predisposes them against such unethical behavior (De Bock & Van Kenhove 2010; Fooks et al. 2012). Indeed, perceived justification for bribery is positively related both to people's intentions to engage in bribery and positive attitudes towards bribery (Powpaka 2002).

Fourth and finally, eliciting participants' perceptions that accepting the bribe would be a mistake directly measures their anticipated regret over accepting the bribe. Regret aversion is a primary driver of aversion to risks and unethical behavior (Loomes & Sugden 1982). Higher anticipated regret moderates ethical decision-making (Pletti et al. 2016) and is positively associated with workplace rule-following (Chen et al. 2024). These four components of the WBS scale are thus all likely positively related to bribery susceptibility. This is true regardless of whether these components are measured for people's susceptibility to bribery offering or acceptance, and regardless of whether they are measured in private-to-private or public bribery settings.

Though this bribery scale is empirically validated for an offer-centric public bribery setting in Weißmüller & De Waele (2022) and De Waele et al. (2021), we show here that our WBS scale is an equivalently valid instrument for measuring the willingness to accept a bribe. Table 1 displays the item-level validation measures of the Likert scale. These measures include

correlation coefficients *r,* Kaiser-Meyer-Olkin (KMO) factor adequacy measures (Kaiser 1970), and uniqueness coefficients constructed from the residual variance on factor loadings.

[Insert Table 1 about here]

All measures reported in Table 1 (and the rest of this subsection) concern the final sample (that is, Dutch native speakers who pass all attention checks). One can see simply by inspection that the Likert items are quite inter-correlated. The KMO measures also demonstrate a meritorious degree of factor adequacy for each of the Likert items. No Likert item appears to be particularly unique (though likelihood and justification appear somewhat more unique as a group than comfort and mistake). We additionally calculate two more statistics across all four of the items. First, Bartlett's (1951) test for sphericity yields a chi-squared statistic of 2,768.44 ($df = 6$, $p < 0.001$). This test evaluates whether the correlation matrix is significantly different from an identity matrix. The significant result indicates high inter-correlation between the variables in the correlation matrix. Second, Cronbach's (1951) alpha is 0.891, revealing sufficiently high internal consistency of the WBS measure.

In addition to providing these summary statistics, we also verify the internal consistency of the WBS measure using both exploratory and confirmatory factor analyses. First, as advocated by Schumacker & Lomax (2010), we begin by randomly splitting the final sample with non-missing data for all covariates, reserving one half of this sample for exploratory factor analysis (EFA) and the other half of the sample for confirmatory factor analysis (CFA). We then perform EFA (specifically principal factor analysis) using the first half of the sample. This procedure results in only one factor being retained, as the eigenvalues for all remaining factors are negative. These results are displayed in Table 2.

[Insert Table 2 about here]

We then utilize the second half of the sample for CFA to confirm that all four components of the WBS measure cleanly map onto one latent construct using a cross-validation

20

sample. Specifically, we construct a structural equation model where one construct (WBS) is mapped onto each of the four components of the WBS measure. This model shows that all four components of the WBS measure are highly significantly positively related to the same underlying WBS construct. We display these results in Table 1. Given all tests, we conclude that the WBS measure utilized here is internally consistent, and hence is a reliable and valid measure.

Other Measures

All Likert scales throughout this experiment are standardized to five-point scales (in accordance with our WBS measure) unless derived from other works that make use of scales with a greater number of points; these exceptions are specially noted. We measure FLP through a two-item Likert measure. One of these items captures participants' self-assessed proficiency with **using** English and the other item captures their proficiency with **understanding** English. In addition to FLP, we also elicit and control for foreign language anxiety (FLA), which is expected to exhibit a similar (inverse) moderating effect on the moral FLE (inverse because higher FLA should in principle yield a higher, not lower, moral FLE). FLA is measured through a ten-item Likert measure, adapted from a much longer scale from the linguistics literature (Gargalianou et al. 2016). These ten items capture various aspects of FL usage that may induce anxiety.

We employ four further categories of control variables in this paper's analysis.[7] The first category consists of demographic measures of gender,[8] age (proxied by birth year), and education level (parameterized as an indicator variable which equals one if the student is a second-year student and zero if they are a first-year student). Gender and age are collected due to their well-studied impacts on ethical decision-making (see, for example, Peterson et al. 2001), and education level is included to control for any business ethics training learned in the first year of study for any of our student participants. All demographic variables considered

here control for the potential impact of differential socialization on ethical behavior, and are shown to be associated with bribery propensity in prior literature (Swamy et al. 2001; Kaufmann et al. 2008; Alatas et al. 2009; Ramdani & van Witteloostuijn 2012; Gatti et al. 2013).

The second group of covariates consists of work-related variables, including years of work experience, public sector employment status, and private sector employment status. No individual in our final sample is employed in both the public and private sectors, though roughly a quarter are employed in neither (see Table 3). Such employment-related variables are long found to impact ethical decision-making in some cases (Ford & Richardson 1994). This is likely because being employed for longer amounts of time (or at all) yields increased social identification with one's work environment and may thus change attitudes about work-related ethical practices, where the private and public sector contexts provide different socialization frames (Weißmüller et al. 2023).

The third group of control variables consists of attitude-related covariates often found to affect ethical decision-making. First, we measure social value orientation (SVO) using the validated unincentivized revealed preference mechanism from Bogaert et al. (2012) in which participants are asked to split hypothetical money between themselves and a stranger. For each of ten splits, participants have the choice to arrange the split to maximize their total earnings (self-maximizing), the difference between their earnings and those of the stranger (competitive), or the total earnings that both players earn (prosocial) across ten choice exercises. We measure SVO in two dimensions by counting both the number of prosocial and the number of competitive choices that each participant makes. SVO is a well-documented moderator for decision-making processes with ethical components (see Gärling et al. 2003, van Dijk & De Cremer 2006; van Prooijen et al. 2008). This is theoretically expectable since individuals with higher levels of SVO hold higher valuations of social welfare *ceteris paribus*,

and are thus expected to behave in a more prosocial manner regardless of treatment exposure. Second, we measure public service motivation (PSM) using the internationally-validated 12-item seven-point Likert scale by Kim et al. (2013). PSM measures individuals' motivation to promote the public good, capturing the dimensions of attraction to public policy making, commitment to the public interest, self-sacrifice, and compassion, and is thus an expectant moderator of integrity-related workplace attitudes and behaviors. Such moderation is indeed frequently empirically found in the public administration literature (see, for instance, Potipiroon & Faerman 2016; Park & Lee 2020).

The fourth and final set of control variables adjusts for components of the experimental design. Participants are randomly assigned to only two of the three vignettes. While this helps control for experimenter demand bias, this design choice raises the possibility that the specific pair of assigned vignettes may influence individuals' responses to the vignettes – participants' judgments of the bribery scenarios may be subject to framing effects induced when comparing the acceptability of taking the bribe between the two vignettes that are presented. For instance, individuals may report lower/higher WBS for the inspector vignette when compared to the negotiator vignette than they would if the inspector vignette is instead compared to the whistleblower vignette. The assigned vignettes are viewed in random order. While we do not collect information on the order in which participants view the assigned vignettes, we naturally have data on which two vignettes each participant views. We thus control for the specific pair of vignettes each participant views with indicator variables for viewing the inspector and negotiator vignettes as a pair ("Inspector-negotiator pair") and viewing the inspector and whistleblower vignettes as a pair ("Inspector-whistleblower pair"). Summary statistics for included covariates in the final sample are displayed in Table 3.

[Insert Table 3 about here]

Descriptive Analysis

Table 4 reports a *post hoc* analysis comparing participants' susceptibility to, and perceived realism of, the acceptance-centric workplace bribery vignettes employed in this study to those of the offer-centric education bribery vignettes employed in De Waele et al. (2021). As aforementioned, our WBS measure is inspired by the bribery measures from De Waele et al. (2021) and Weißmüller & De Waele (2022). Although Weißmüller & De Waele (2022) do not report summary statistics on WTB, De Waele et al. (2021) do report such summary statistics, and thus we can directly compare our bribery susceptibility measure to the measure elicited in De Waele et al. (2021). This comparison is made even more useful by the fact that the samples in De Waele et al. (2021) are very similar to our final sample. Across all samples, all participants are university students in business, economics, or social sciences. De Waele et al. (2021) even report means and standard deviations of their measures for samples of students at Dutch and Belgian universities who were provided Dutch-language vignettes, permitting within-language and within-country comparisons.

[Insert Table 4 about here]

Table 4 shows that participants report significantly higher susceptibility to our acceptance-centric workplace bribery vignettes than to the offering-centric education vignettes in De Waele et al. (2021). These differences range from small to moderate in Cohen's *d* effect size terms. Across all comparison samples from De Waele et al. (2021), bribery susceptibility is .398*d* to .621*d* higher for our vignettes than for those in De Waele et al. (2021). These differences are not explained by our FL treatment, and they are unlikely to be explained by differences in sample characteristics. Though limiting our sample only to observations facing the NL treatment condition slightly reduces these differences, bribery susceptibility is still .398*d* to .575*d* higher for our vignettes than for those in De Waele et al. (2021) amongst participants facing bribery vignettes in their NL. In fact, when comparing the samples of Dutch

24

university students facing Dutch-language vignettes, those participants in our sample exhibit .501d higher bribery susceptibility than those in De Waele et al. (2021).

These results suggest that our participants exhibit higher bribery susceptibility because they face vignettes about accepting a private-to-private bribe rather than vignettes about offering a public bribe. This is in line with theoretical expectations. As previously discussed, people are likely more susceptible to accepting than to offering bribes and are likely more susceptible to private-to-private bribery than to public bribery. Disentangling which of these two mechanisms explains the higher bribery susceptibility found in our sample is a fruitful topic for future research.

Though most of our manipulation checks are unfortunately absent from De Waele et al. (2021), one that is present in De Waele et al. (2021) is vignette realism. De Waele et al. (2021) measure this on a four-point Likert scale, while we measure realism on a five-point Likert scale. To ensure that the two measures are comparable, we simply scale the reported means and standard deviations of vignette realism from De Waele et al. (2021) by a factor of 1.25. This ensures that the two realism measures have the same range.

Table 4 shows that participants find our acceptance-centric business bribery vignettes to be significantly more realistic than the offering-centric education bribery vignettes in De Waele et al. (2021). These differences are more modest than those concerning bribery susceptibility; in Cohen's d terms, they are small but still meaningful, ranging from .257d to .481d. As with bribery susceptibility, these differences are not explained by our FL treatment or by sample characteristics. In fact, the second-largest observed difference in perceived vignette realism is between our sample facing the NL treatment condition and the Dutch student sample in De Waele et al. (2021), sitting at .470d.

These results show that participants perceive our vignettes to be much more realistic than bribery vignettes in the prior literature. Our vignettes are also perceived as more realistic

than not on average, with a mean realism score exceeding three (see Table 3). Further, these

vignettes are certainly more realistic than the highly artificial footbridge and trolley dilemma

vignettes that dominate prior moral FLE experiments (for example, see Costa et al. 2014;

Geipel et al. 2015; Hayakawa et al. 2017; Stankovic et al. 2022). This implies that our

experiment offers a unique opportunity to experimentally examine the moral FLE in a setting

that is far less artificial than those in which the moral FLE is examined in prior research.

Results

Main Results

The main regression results on the WBS measure are shown in Table 5. The first row ("FLE") displays the marginal FLE on WBS, while the remainder of the rows show the full results of the regression model. For the models without an interaction effect (Models 1 and 2), the marginal effect of the English treatment is simply equivalent to the coefficient on the English treatment dummy (which is why the estimates for 'English' and 'WBS' are identical in Models 1 and 2). However, in the models with an interaction effect (Models 3 and 4), the main effect is the coefficient on English × FLP times the average FLP in the sample, plus the coefficient on English. This is automatically calculated in Stata as a post-estimation using the 'margins' suite and appended to the top of the table. Model 4 is the primary model of interest for two reasons. First, the full control specification corrects for any chance imbalances in the specified covariates. As can be seen by the progression of the FLE estimates as more covariates are incorporated, these imbalances are inflating the initial FLE estimate, and this inflation is corrected in the fully specified model. Second, the variance of the FLE treatment effect estimator in Model 4 is 8% lower than that of Model 1. This implies that incorporating the additional controls is improving the efficiency of the FLE estimator by more accurately capturing variance in WBS's data-generating process.

Hence, we primarily discuss the results from Model 4. The standard errors in Model 4 imply that the treatment effect estimator for the FLE in Model 4 is precise enough to detect small effect sizes at a 0.5% significance level.[9] Even without the added predictive power of our covariates, due to the high number of observations (616 in the NL treatment, 596 in the FL treatment), a simple *t*-test of WBS means between treatment groups in our sample can detect a small effect size of $d = 0.2$ with 96.1% power at a 5% significance level and 81.8% power at a 0.5% significance level. However, all models yield the same conclusions under an NHST

framework for reasonable significance levels, and imprecision cannot explain insignificant findings in any of the models in Table 5.

[Insert Table 5 about here]

The results show that FL usage does not significantly impact WBS in our sample. While we correctly hypothesize the positive sign of the FLE on WBS, which consistently emerges across all model specifications, the observed FLE is quite small ($0.062 < d < 0.086$) across all models and is insignificant for reasonable significance thresholds. We thus fail to find support for H1. These findings are visible in the WBS averages by language and vignette, as visualized in Figure 2 with 95% confidence intervals.

[Insert Figure 2 about here]

Additionally, we do not find that FLP significantly moderates the FLE, although the sign is again consistent with the hypotheses across all specifications. That is, while the point estimate on the interaction term between the FL treatment and FLP is negative (as predicted) and quite large (with a one-point increase in FLP being associated with a reduction in the FLE roughly equivalent to the FLE point estimate itself), this estimate is very noisy ($t = 1.079$). We visualize this in Figure 3. The dots and confidence bands in Figure 3 respectively depict predicted WBS levels and simple 95% confidence intervals by FLP level and language condition, where WBS predictions are generated from the regression of Model 4 in Table 5. It is clearly visible that there is no level of FLP at which the FLE becomes significant (even before making multiple-hypothesis corrections). We thus find no significant evidence for H2.

[Insert Figure 3 about here]

We now assess the certainty of our null findings via conditional equivalence testing (Campbell & Gustafson 2018). Since a statistically insignificant regression coefficient does not necessarily imply no practical treatment effect on WBS, we use a two one-sided tests (TOST) procedure to assess, for some effect size d, whether the regression coefficients on the FLE main

28

and the English-FLP interaction effects are bounded between -d and d using two one-sided tests (Lakens et al. 2018). To our knowledge, this is the first instance where TOST or any comparable equivalence testing procedure has been used in the business ethics literature. We use Cohen's (1988) effect size benchmarks of $d = 0.2$ for small effect sizes, $d = 0.5$ for medium effect sizes, and $d = 0.8$ for large effect sizes.[10] Equivalence testing results using estimates from Model 4 of Table 5 are shown in Table 6. We cannot rule out small effect sizes for the English-FLP interaction effect estimator, and thus conclude that our examination of the moderating effect of FLP on the moral FLE is inconclusive. However, results suggest that the FLE on WBS is significantly bounded below small effect sizes.[11] Results demonstrate with high certainty that all treatment effects are bounded beneath all effect sizes at and above the medium level. Results thus suggest that any FLE which may exist in our data is **smaller than small**. Our analysis thus ultimately provides suggestive evidence against H1, and inconclusive evidence on H2.

[Insert Table 6 about here]

Robustness Checks

We conduct several *post hoc* checks to assess the robustness of our null findings. First, we repeat the analysis from this section with an alternative parameterization of our dependent variable. Specifically, we use the first half of the sample preserved for EFA to predict values of the latent WBS construct for the whole sample using a regression scoring method (specifically Stata's 'predict' post-estimation command after 'factor'). We then repeat the analyses in Tables 5 and 6 using this predicted WBS variable as the outcome. Predicted WBS is effectively a slightly weighted mean of participants' responses to the four components of our initial WBS measure, rather than a simple mean. Appendix C displays the results of this analysis in Tables A1 and A2. These analyses continue to provide evidence against H1, and also provide suggestive evidence against H2. Table A1's FLE estimates and English x FLP

29

estimates are both smaller and more precisely estimated than those from Table 5, resulting in significant equivalence testing *t*-statistics in Table A2 for both the FLE and the English x FLP moderation effect. This indicates that the predicted WBS parameterization yields suggestive evidence that the moderating effect of FLP on the FLE is bounded beneath small effect size benchmarks. However, given the insignificant equivalence testing results for the English x FLP moderating effect in Table 6, we do not view this evidence as particularly robust.

Second, we examine the robustness of our findings across vignettes by disaggregating our sample by vignette and re-doing the estimation procedure for Model 4 in Tables 5 and 6 within each vignette-level partition. Appendix D displays these results in Tables A3 and A4. Table A3 shows that the main FLE estimates are not statistically significantly different from zero in any of the three vignettes. However, Table A4 shows that statistically significant evidence for the FLE's practical equivalence to zero disappears for the inspector and whistleblower vignettes, only remaining (at the 5% level) for the negotiator vignette, which exhibits the smallest FLE of the three vignettes. This is expectable; partitioning the data in this fashion drops roughly two thirds of the sample, which forces a steep drop in power. The standard errors of the vignette-level FLE estimates in Table A4 exceed that of the pooled FLE estimate by 47-90%. This implies that the vignette-level FLE estimates' minimal detectable effect sizes for given significance and power levels exceed that of the pooled FLE estimate by at least that magnitude, in turn implying that the vignette-level FLE estimates exhibit much less power than the pooled FLE estimate (Bloom 1995). However, for all three vignettes, the point estimates for the FLE are well beneath a small effect size of .214 (see Table 6), and no outlier vignette exists where the FLE is significantly different from zero.

Additionally, Table A3 shows that like the pooled English x FLP interaction effects in Table 5, the vignette-level English x FLP interaction effects are large but quite noisy. Only one of the three is statistically significantly greater than zero (specifically that for the whistleblower

30

vignette), and this significance conclusion would not survive a simple Bonferroni-Holm correction for multiple hypothesis testing (Holm 1979). Similarly to the equivalence testing results for the pooled English x FLP interaction effect, the equivalence testing results for all three vignette-level English x FLP interaction effects produce no significant evidence that the interaction effect is practically equivalent to zero. These results for the moderating effect of FLP on FLEs are thus ultimately inconclusive.

Manipulation Checks

The observed treatment effect on the negotiator vignette from Table 5 is remarkable for its large effect size ($d = 1.520$ in Model 4) and strong significance ($t = 19.441$ in Model 4). Such an effect is worthy of *post hoc* examination. Why does this effect arise? A clue may exist in our manipulation checks, the results of which are displayed in Table 7.

[Insert Table 7 about here]

These tests show significant evidence that the different vignette settings successfully manipulated participants' perceptions of the bribery scenarios. We can assess whether one of the constructs measured in our manipulation checks is consistent with the observed vignette effects on WBS by examining whether the conclusions arising from some monotonic transformation of its regression estimates in Table 7 match those arising from the vignette effect estimates found in Table 5. This rules out both realism and benefit to the briber as consistent explanations.[12] However, a monotonic transformation of the results for harm to the organization is consistent with the results in Table 5. The signs of the negotiator vignette and whistleblower vignette coefficients in Table 7 are exactly opposite to those of Table 5. The negotiator vignette also induces a significant difference in both WBS and perceptions of harm to the organization. In contrast, there is no significant difference between the inspector and whistleblower vignettes for either WBS or perceived harm to the organization. This suggests a possibility that individuals may be more willing to accept bribes that they perceive as being

31

less harmful to the goals of their organization. However, our results are not conclusive on this matter and more research is needed to establish any such effect.

Discussion

This study contributes experimental evidence to the literature on bribery by adopting a realistic vignettes design. Different from what is usually seen in the literature on bribery's microfoundations (and the business ethics literature more broadly), we focus on the willingness to accept rather than to offer bribes in a private sector-only context, shifting focus to the demand side of bribery. We find suggestive evidence against the existence of an FLE on WBS and find inconclusive evidence regarding the moderating effect of FLP on this FLE. It is important to report such null findings, particularly when theory and prior literature would strongly predict otherwise, as is the case for our two moral FLE hypotheses. We also offer empirical evidence suggesting a fruitful future route of research, with *post hoc* analyses implying a possibility that individuals may be more willing to accept bribes that they perceive as being less harmful to the goals of their organization. Future research is needed before such a conclusion is considered to be confirmed.

The null FLE found in this study is striking in part because the vignette conditions we construct are favorable to significant FLEs. There is a paucity of regulatory legislation or enforcement against private-to-private bribery, which both arises from and induces weak prescriptive norms against accepting the bribes in our vignettes (see Argandoña 2003; Argandoña 2005; Gopinath 2008). Additionally, the high perceived descriptive norms in favor of private-to-private bribery (Jaakson et al. 2019) together with the temporary spike in such perceived descriptive norms when facing a bribe offer combine to imply that participants are quite likely to engage in 'no-big-deal' heuristic thinking for our set of vignettes. Because the theoretical mechanism behind the FLE relies on a switch between heuristic System 1 thinking to deliberative System 2 thinking (see Keysar et al. 2012; Greene 2014; Geipel et al. 2015; Białek et al. 2019), these conditions give FLEs a great chance to arise. Because our vignettes induce more participants to engage in heuristic thinking at baseline, more participants can be

'switched' to System 2 thinking via FL usage. However, we observe suggestive evidence against the existence of such an effect. This increases the chance that our results will replicate beyond this study and extend into real-world settings, which may be less favorable to the existence of significant FLEs.

This study contributes to the literature in at least four ways. First, we expand the tiny business ethics literature on the moral FLE (Pan & Patel 2018; Sugahara et al. 2023). We specifically examine this effect on an important form of corporate corruption (that is, private sector bribery). The absence of FLEs in this setting implies that managerial decisions on language accommodation in international business are unlikely to invoke changes in organizational corruption or ethical behavior, at least in settings with high FLP.

Second, we show that the moral FLE may not be generalizable to applied business settings. While a positive FLE on utilitarian behavior and a related attenuative FLP moderating effect are often found in more artificial experimental settings (Circi et al. 2021; Stankovic et al. 2022), we examine a new setting that we show is perceived as significantly more realistic than vignettes in prior experimental bribery research, and is certainly more realistic than the trolley/footbridge dilemma vignettes employed in prior moral FLE research. In a more realistic setting with applied business scenarios, we find suggestive evidence against the existence of FLEs, and inconclusive evidence regarding the attenuative moderating effect of FLP on the FLE.

Third, we investigate unique and useful bribery settings. Our private sector bribery vignettes extend experimental investigation of bribery determinants to the private sector-only context. Our experiment additionally introduces a focus on the often-ignored demand side of bribery by investigating determinants of bribery acceptance rather than bribery offering. This is critical to understanding how bribery manifests, as both bribery offering and acceptance are necessary for bribery to take place.

34

Fourth and finally, we introduce equivalence testing to business ethics' standard hypothesis testing toolkit. This study provides business ethics researchers with a template on how to demonstrate significant evidence that treatment effects and/or variable associations are practically equivalent to zero (such as our FLE estimate), rather than relying on the flawed inference that all effects/associations that are not significantly different from zero are 'statistically indistinguishable' from zero (Altman & Bland 1995). Note that estimates may be neither significantly different from zero nor significantly equivalent to zero. In cases where this occurs (for instance, our FLE x FLP interaction effect), the researcher should conclude that the evidence on this effect/association is inconclusive, as there is not strong evidence either against or for the null hypothesis of no effect/association. Adopting these techniques will make null results more credible and easier to publish, an essential first step to combating publication bias (Kepes et al. 2014). Because any statistical relationship may be practically equivalent to zero, this is a testing procedure that all business ethics scholars should be prepared to employ for any empirical study they undertake.

Limitations and Future Research Directions

One potential explanation for why we may observe so small an effect size for the FLE might be the specific pair of languages selected. English is extremely ingratiated into Dutch society, with 90% of Dutch citizens being conversationally proficient in English (Eurobarometer 2012). English is ubiquitous throughout Dutch media, education, and society (Edwards 2016). Dutch and English are also remarkably close languages – English is lexicostatistically closer to Dutch than any other major Indo-European language (Dyen et al. 1992; Ginsburgh & Weber 2016). The data reflects these factors. Figure 4 displays a histogram of FLP levels for native Dutch-speaking study participants who passed all attention checks.

[Insert Figure 4 about here]

Nearly 72 percent of our participants rate the average of their active and passive proficiency in English at a four out of five or higher. The invariantly high levels of English language proficiency and usage found in the Netherlands (and our sample) may attenuate observed FLEs when comparing English and Dutch (Stankovic et al. 2022). Furthermore, the high level of FLP in our sample explains the wide confidence bands near the bottom of the FLP distribution in Figure 3. Our sample is thus not well-powered to discover heterogeneous FLEs near the bottom of the FLP distribution, which is likely true for most if not all Dutch samples that examine an English FLE. Future extensions of our study design are encouraged to use different pairs of languages. However, the selection of these language pairs should account for issues of generalizability and societal relevance. The most societally relevant FLEs are those concerning languages that are interchanged most often. However, in many countries, the languages that are most often interchanged within a given population are likely to be languages that are highly similar and/or languages for which the population possesses a high degree of proficiency.

An alternative research trajectory is thus to examine the FLEs amongst multiple pairs of languages at once. This practice is common in the FLE literature (Circi et al. 2021) and eliminates the concern that the observed presence or absence of an FLE is simply an artefact of one given pair of selected languages. However, this research trajectory generates a conflict between internal and external validity. The common practice amongst FLE studies which examine multiple language pairs is to use a different sample for each language pair examined. While this practice is practical and necessary for some language pairs, the practice also induces potential confounding between language pairs and pre-treatment characteristics. The experimental best practice for preventing such confounding is to examine participants who know three or more languages, and randomize the assigned pair of languages amongst that sample. However, such individuals may not only be very difficult to find in sufficiently large

numbers for specific trios of languages, but are also unlikely to be representative of the general population – for example, only 35% of the European population speaks three or more languages (Eurobarometer 2012). FLEs observed amongst such participants may therefore face generalizability challenges.

Another limitation of our study is the lack of an incentivized task. This is not uncommon in the moral FLE literature. None of the 38 examined moral FLE experiments in the Circi et al. (2021) meta-analysis used task-related incentives, though it is certainly possible to experimentally analyze FLEs in incentivized contexts (see, for instance, Bereby-Meyer et al. 2020; Alempaki et al. 2021). It is potentially fruitful to examine whether observed FLEs differ in incentivized and unincentivized conditions. However, we believe that our conclusion here is unlikely to change under different incentivization schemes since incentivized FLE studies seem to exhibit attenuated treatment effects. For example, the risk FLE experiments in Circi et al. (2021) that use task-related incentives exhibit smaller treatment effect sizes and test statistics than those found in the risk FLE experiments which do not employ such incentives.[13] Such an attenuating effect would only strengthen our finding of a moral FLE bounded below small effect sizes.

Our *post hoc* analyses also suggest several bribery microfoundations which may be fruitful to investigate in future experiments. First, Table 4 shows that bribery susceptibility is higher in our vignettes, which are acceptance-centric and concern private-to-private bribery, than for the vignettes in De Waele et al. (2021), which are offering-centric and concern public bribery (specifically in an education context). These results are consistent both with theoretical expectations for the differences in susceptibility to bribery acceptance and bribery offering, as well as those for the differences in susceptibility to private-to-private and public bribery. Future experimental research could explore both conditions in a 2x2 factorial design. Assuming that the main effect of these conditions on bribery susceptibility replicates, this design would permit

37

researchers to examine whether each of the two factors – offering vs. acceptance and public vs. private-to-private bribery – impacts bribery susceptibility independently, or whether they are required in consort to yield the same effect. Second, our results in Table 7 show that the FLEs and vignette effects on WBS, as well as those effects on participants' perceptions of the proposed bribery's harm to their hypothetical business organization, share exactly opposite signs as well as the same significance conclusions. This yields a theoretically sensible hypothesis: people may be more susceptible to bribes that are less harmful (or more beneficial) to their business organization (either in perception or in fact). While a future experiment is needed to confirm or deny this hypothesis, one could reasonably vary the stakes of bribery harm (and/or benefit) to the business organization as a treatment to discern whether WBS is impacted by these stakes.

Conclusion

Drawing lessons for practice from very simplified experimental settings should be done with much caution. This is not different in our case, notwithstanding the realism of our vignettes and the power of our sample. But would our null findings survive further scrutiny in future research, also outside the lab context, then two key takeaways, one for researchers and one for managerial practice, can be derived from our empirical findings. First, researchers should rest assured that if experimental participants are sufficiently proficient in two languages, the choice of which language is displayed to such participants will not have an outsized impact on treatment effects for vignettes in business ethics. Although it is still useful to remain transparent about the language in which experimental participants face survey stimuli, in research settings where it is reasonably certain that experimental participants are highly proficient in a *lingua franca*, business ethics researchers should feel comfortable designing experimental stimuli in the *lingua franca*. This can ensure that all members of research teams

38

from diverse linguistic backgrounds can easily read, understand, and contribute to survey materials, and may help improve the power of such experimental research by reducing the attrition of participants who speak the *lingua franca*, but not the native language. Second, businesses should not allow concerns of potential negative moral FLEs to impact decisions on internationalization, as there is suggestive evidence that this effect does not arise in applied business settings.

Tables

Table 1: Item-Level Validation of the WBS Measure

	Likelihood	Comfort	Justification	Mistake
$r_{\text{Likelihood}}$	1.000			
r_{Comfort}	0.660	1.000		
$r_{\text{Justification}}$	0.658	0.654	1.000	
r_{Mistake}	0.654	0.667	0.743	1.000
SEM coefficient	1.000	0.919	1.105	1.058
		(0.047)	(0.050)	(0.051)
KMO	0.860	0.857	0.816	0.813
Uniqueness	0.390	0.383	0.313	0.306

Note: All correlation coefficients r are significant with $p < 0.001$. Kaiser-Meyer-Olkin factor adequacy measures (KMO) are displayed below correlation coefficients and SEM coefficients alongside variable uniqueness measures, the last of which are calculated using the residual variance of factor loadings. SEM coefficients display the coefficients from a structural equation model where construct 'WBS' is modeled with paths to each of the four components of the WBS measure. The coefficient on the first component (Likelihood) is constrained to be equal to one by construction. For the other three components, standard errors are reported under their respective SEM coefficients in parentheses. The SEM estimation is conducted with 602 observations.

Table 2: Exploratory Factor Analysis Results for the WBS Measure

	Factor 1	Factor 2	Factor 3	Factor 4
Eigenvalue	2.642	-0.040	-0.100	-0.118
Change in eigenvalue	2.681	0.061	0.017	

Note: This table reports eigenvalues from an EFA (specifically a principal factor analysis) conducted upon a randomly split sample of

participants' responses for each of the components of the WBS measure. Eigenvalues and changes in eigenvalues (from a higher factor to the

next lowest factor) are reported. This estimation is conducted with 604 observations.

Table 3: Summary Statistics

Variable	N	Mean	Median	SD	Min	Max
WBS	1,212	2.52	2.25	1.07	1	5
FLP	1,212	4.05	4	0.72	1	5
FLA	1,212	2.35	2.2	0.84	1	4.90
Birth year	1,212	2002.42	2003	1.45	1993	2005
Female	1,206	0.37	0	0.48	0	1
Second-year student	1,212	0.47	0	0.50	0	1
Years of work experience	1,212	3.67	4	1.88	0	14
Employed in public sector	1,212	0.31	0	0.46	0	1
Employed in private sector	1,212	0.44	0	0.50	0	1
SVO prosocial choices	1,212	4.86	6	3.92	0	9
SVO competitive choices	1,212	0.17	0	0.99	0	9
PSM	1,212	4.14	4.08	0.89	1.33	6.58
Inspector-negotiator pair	1,212	0.34	0	0.47	0	1
Inspector-whistleblower pair	1,212	0.32	0	0.47	0	1
Realism	1,212	3.08	3	1.09	1	5
Benefit to the briber	1,212	3.85	4	1.11	1	5
Harm to the organization	1,212	3.61	4	1.22	1	5

Note: Our observation count is double the number of participants as WBS is measured for each participant across two vignettes.

Table 4: Mean Differences in Bribery Susceptibility Measures and Vignette Realism Judgments between WBS and WTB

	(1)	(2)	(3)	(4)	(5)	(6)
Bribery susceptibility	.580	.535	.470	.425	.650	.605
	(.082)	(.088)	(.079)	(.085)	(.077)	(.083)
	{.550d}	{.501d}	{.446d}	{.398d}	{.621d}	{.575d}
Vignette realism	.517	.507	.293	.282	.455	.445
	(.083)	(.089)	(.081)	(.088)	(.079)	(.084)
	{.481d}	{.470d}	{.269d}	{.257d}	{.425d}	{.415d}
Comparison sample	Dutch	Dutch	German	German	Belgian	Belgian
NL restriction		x		x		x
N, our sample	1212	616	1212	616	1212	616
N, comparison sample	193	193	211	211	220	220

Note: This table displays results from an immediate t-test of differences in means of the row variable between the column sample in this study and the column comparison sample from De Waele et al. (2021), based on the reported sample sizes, means, and standard deviations of that row variable provided for that comparison sample in De Waele et al. (2021). Bribery susceptibility is WBS in this study and WTB in De Waele et al. (2021). Beneath differences in means, pooled standard errors are displayed in parentheses and Cohen's d effect sizes are displayed in curled brackets. For columns where the NL restriction is indicated, the difference in means is reported only for observations in our final sample facing the Dutch-language treatment.

Table 5: Main Regressions

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
FLE	.092	.084	.083	.066
	(.058)	(.054)	(.054)	(.053)
English	.092	.084	.499	.415
	(.058)	(.054)	(.324)	(.320)
Negotiator vignette		1.298	1.300	1.325
		(.063)	(.063)	(.068)
Whistleblower vignette		-.016	-.013	-.069
		(.051)	(.051)	(.061)
English x FLP			-.103	-.086
			(.079)	(.079)
FLP			.043	.010
			(.059)	(.068)
FLA				-.024
				(.043)
Birth year				-.025
				(.020)
Female				-.136
				(.059)
Second-year student				-.024
				(.059)
Years of work experience				-.008
				(.016)
Employed in public sector				.162
				(.081)
Employed in private sector				.082
				(.074)
SVO prosocial choices				-.015
				(.007)
SVO competitive choices				.004
				(.027)
PSM				-.062
				(.035)
Inspector-negotiator pair				-.087
				(.076)
Inspector-whistleblower pair				.045
				(.070)
Constant	2.475	2.042	1.865	52.14
	(.041)	(.047)	(.248)	(40.921)
Obs.	1,212	1,212	1,212	1,206
F	2.537	178	107	35.92
R^2	.002	.337	.339	.360
Adj. R^2	.001	.336	.336	.351

44

Note: For each model, estimate 'FLE' reflects the marginal effect of the English language treatment on WBS, calculated using the coefficient 'English' and, in Models 3 and 4, the 'English x FLP' coefficient multiplied by the expected value of FLP. The remaining estimates are simply regression coefficients. For all estimates, cluster-robust standard errors are in parentheses.
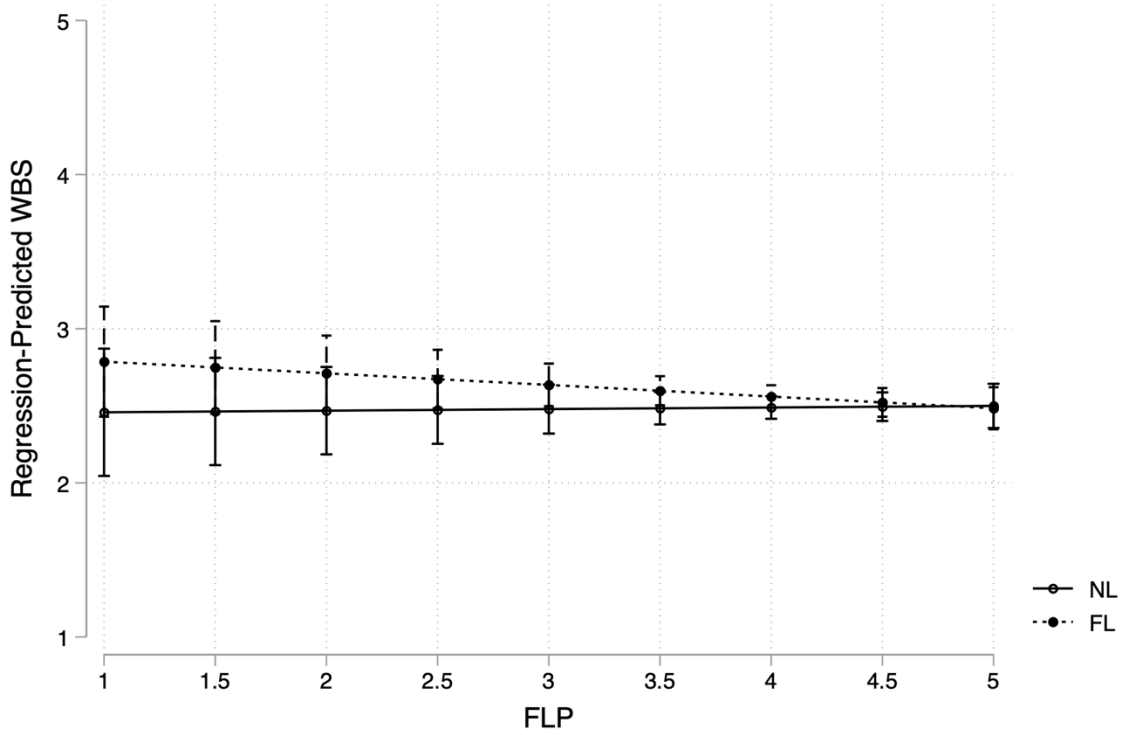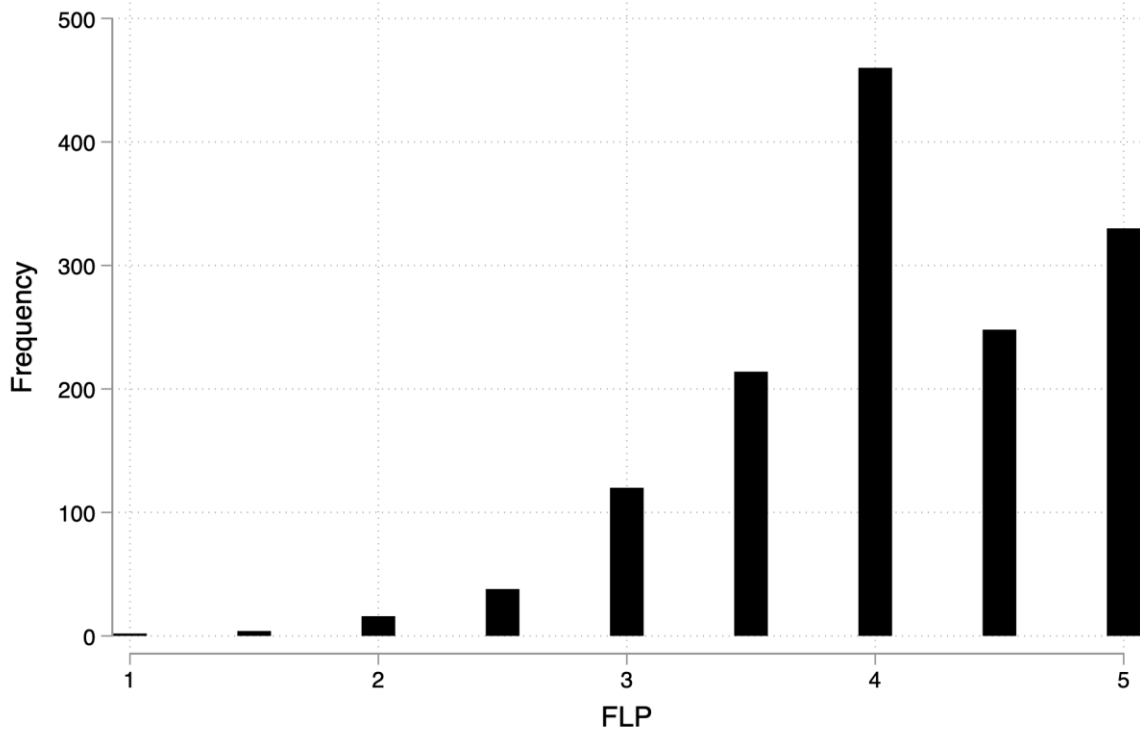
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 6: Equivalence Testing Results

	Small Effect Size	Medium Effect Size	Large Effect Size
Cohen's *d*	.2	.5	.8
Regression coefficient	.214	.534	.855
FLE *t*-statistic	-2.796	-8.855	-14.914
	(.005)	(< .001)	(< .001)
English x FLP *t*-statistic	1.627	5.708	9.790
	(.052)	(< .001)	(< .001)

Note: Cohen's *d* values are provided directly above the regression coefficients corresponding to their respective Cohen's *d* value. The

equivalence tests here assess whether the observed FLE is significantly bounded within the range [-*d*, *d*]. *t*-statistics with corresponding one-

sided significance levels in parentheses are calculated using estimates from Model 4 in Table 5. *p*-values are replaced with Bonferroni-Holm

corrected *q*-values to correct the family-wise error rate (Holm 1979; Newson 2010).

Table 7: Manipulation Checks

	(1) Realism	(2) Benefit to the Briber	(3) Harm to Organization
English	.027	-.003	-.111
	(.067)	(.060)	(.068)
Negotiator vignette	-.674	-.946	-1.030
	(.069)	(.069)	(.074)
Whistleblower vignette	-.290	.336	.144
	(.071)	(.065)	(.076)
Constant	3.392	4.064	3.966
	(.065)	(.060)	(.068)
Obs	1,212	1,212	1,212
R²	.065	.241	.188
Adjusted R²	.062	.239	.186

Note: Regression coefficients on each manipulation check variable are displayed with cluster-robust standard errors in parentheses.

Figures

Figure 1: Participant Flow
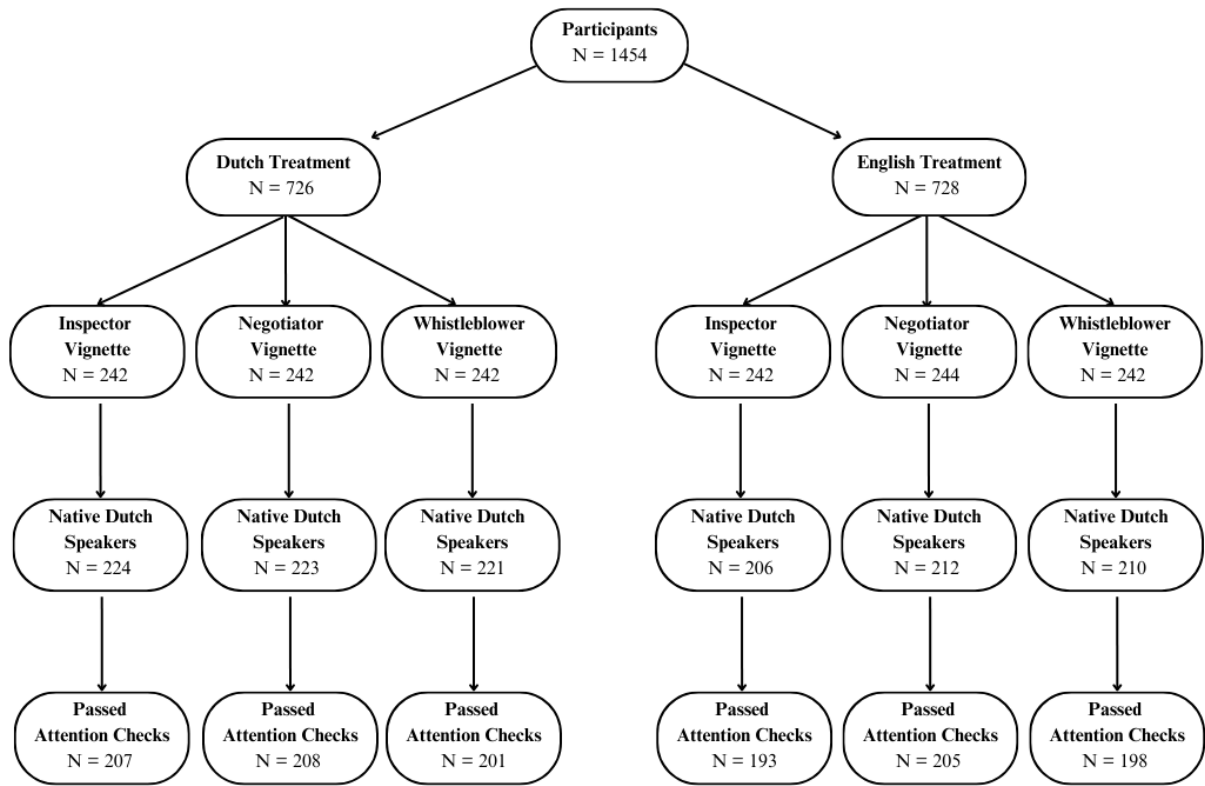
Figure 2: WBS Means by Language and Vignette
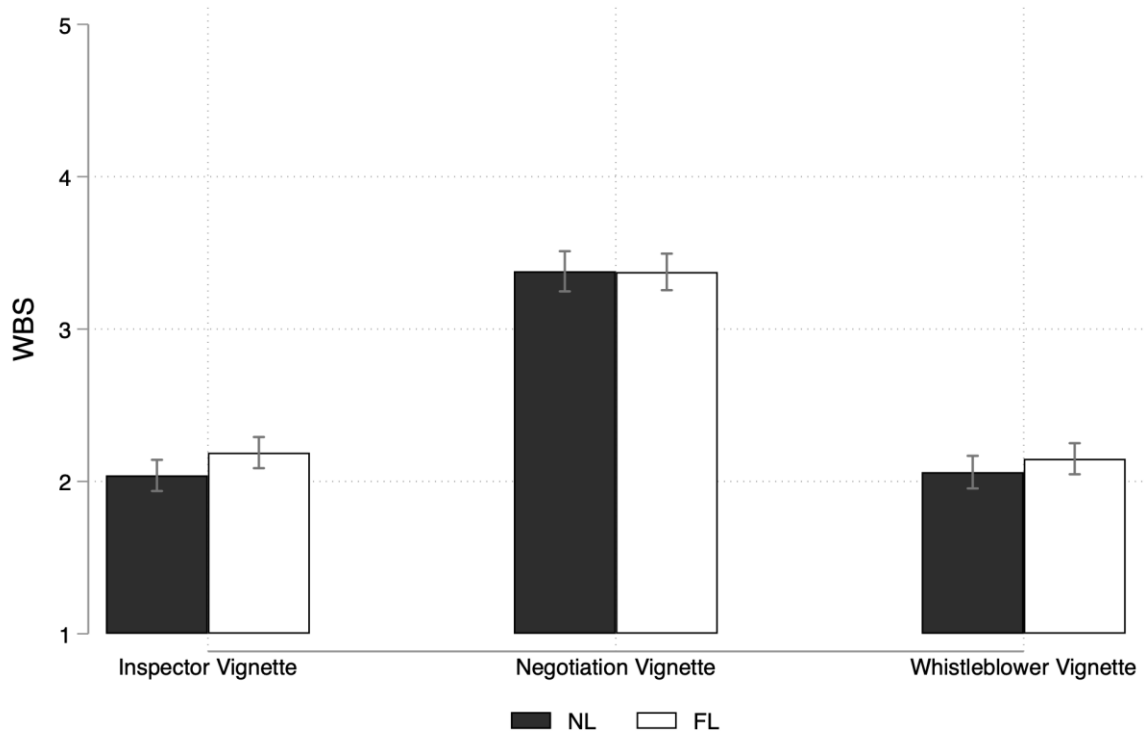
Figure 3: Predicted Margins of WBS by FL Treatment Status and FLP Level

Figure 4: Histogram of FLP in the Final Sample

Figure Legends

Figure 1: Participant counts N represent the total number of **observations**, which is double the number of participants in the top two rows of cells, as WBS is measured for each participant across two vignettes.

Figure 2: Raw WBS means by language condition and vignette are displayed with 95% confidence intervals.

Figure 3: Predicted average WBS levels and 95% confidence intervals by language condition and FLP level are calculated via the Stata 'margins' suite using the coefficient estimates from Model 4 in Table 5.

Figure 4: Frequency counts represent the total number of observations, which is double the number of participants, as WBS is measured across two vignettes for each participant. This fact does not affect relative FLP frequencies.

Appendix A: WBS Measurement

We present here the English-language versions of the questions used to elicit our Likert measures. The questions are presented in the following order. The first question concerns **realism**: "This statement appears realistic." The Likert options are:

1. Totally disagree

2. Rather disagree

3. Neutral

4. Rather agree

5. Totally agree

The remaining questions have no general question format and the specific question statement is provided below. However, the answers do have a general format. The second question concerns **likelihood**. The Likert options are:

1. Very unlikely

2. Rather unlikely

3. Neither likely nor unlikely

4. Rather likely

5. Very likely

The third question concerns **justification**. The Likert options are:

1. Not justified at all

2. Not justified

3. Neither justified nor not justified

4. Justified

5. Very justified

The fourth question concerns **comfort**. The Likert options are:

1. Very uncomfortable

2. Rather uncomfortable

3. Neither comfortable nor uncomfortable

4. Rather comfortable

5. Very comfortable

The fifth question concerns **mistake**, the sixth question concerns **benefit to the briber**, and the seventh and final question concerns **harm to the organization**. The Likert options for all three of these last questions are:

1. Totally disagree

2. Rather disagree

3. Neither agree nor disagree

4. Rather agree

5. Totally agree

The remaining questions asked after each vignette are listed below.

Inspector Vignette

- Likelihood: How likely is it that you would waive the inspection in this case?

- Justification: How justified is it that you would waive the inspection in this case?

- Comfort: How would you feel about waiving the inspection in this case?

- Mistake: I think that waiving the inspection in this case would be a mistake.

- Benefit to the briber: I think that waiving the inspection is beneficial for the owner of the restaurant.

- Harm to the organization: I think that waiving the inspection is adverse for achieving the goals of the organization I work for.

Negotiator Vignette

- Likelihood: How likely is it that you would use the car in this scenario?

- Justification: How justified is it that you use the car in this scenario?

- Comfort: How would you feel about using the car in this scenario?

- Mistake: I think that using the car would be a mistake.

- Benefit to the briber: I think that using the car would be beneficial for the partnering country.

- Harm to the organization: I think that using the car would be adverse for achieving the goals of the organization I work for.

Whistleblower Vignette

- Likelihood: How likely is it that you will withdraw your testimony in this scenario?

- Justification: How justified is it that you withdraw your testimony in this scenario?

- Comfort: How would you feel about withdrawing your testimony in this scenario?

- Mistake: I think that withdrawing my testimony would be a mistake.

- Benefit to the briber: I think that withdrawing my testimony would be beneficial for the suspect.

- Harm to the organization: I think that withdrawing my testimony would be adverse for achieving the goals of the organization I work for.

Appendix B: Vignettes

We present here the English-language versions of the bribery vignettes.

Inspector Vignette

You work as an inspector for a private company that is officially charged with controlling restaurants regarding compliance with the food safety instructions. This company is a private corporation that is profit-oriented. One day, you have to control your favorite restaurant. You know the owner of the restaurant pretty well since you visit this restaurant on a weekly basis. The owner of the restaurant offers you a free meal together with your wife and children if you waive the inspection and give him a maximum score.

Negotiator Vignette

You work as a manager in the international business development unit of a private company and you are abroad negotiating a materials contract from a foreign supplier. As you arrive at your hotel, you find the keys of a very exclusive sports car lying on your bed together with the message: 'For a fruitful collaboration between our companies.' You realize there is no way to bring this car home with you, but it may be nice to drive around for the rest of the trip.

Whistleblower Vignette

You work at an international shipping company. As a whistleblower, you are asked to testify against an important supervisor at your department because he is suspected of having received bribes for warning drug cartels upfront when controls by the anti-drug department were to be carried out. You certainly know that the suspect is guilty and, therefore, deserves to be punished accordingly. However, the day before your testimony, you receive an envelope in your mailbox with the message 'Silence is gold'. The letter further states that you will receive another envelope, containing €10.000 if you decide not to testify.

Appendix C: Predicted WBS Results

Table A1 provides the results of the regressions using the predicted WBS construction discussed in the subsection 'Robustness Checks.' Table A2 subjects the results of Model 4 from Table A1 to equivalence testing in the same fashion as Table 6.

Table A1: Regressions on Predicted WBS

	Model 1	Model 2	Model 3	Model 4
FLE	.090	.083	.083	.067
	(.050)	(.046)	(.046)	(.046)
English	.090	.083	.441	.364
	(.050)	(.046)	(.280)	(.277)
Negotiator vignette		1.134	1.135	1.159
		(.055)	(.055)	(.060)
Whistleblower vignette		-.015	-.013	-.060
		(.044)	(.044)	(.054)
English x FLP			-.088	-.073
			(.068)	(.068)
FLP			.038	.009
			(.051)	(.059)
FLA				-.022
				(.037)
Birth year				-.022
				(.018)
Female				-.115
				(.051)
Second-year student				-.014
				(.051)
Years of work experience				-.008
				(.014)
Employed in public sector				.141
				(.070)
Employed in private sector				.068
				(.064)
SVO prosocial choices				-.012
				(.006)
SVO competitive choices				.006
				(.023)
PSM				-.055
				(.030)
Inspector-negotiator pair				-.072
				(.066)
Inspector-whistleblower pair				.042
				(.061)
Constant	-.021	-.399	-.555	43.361
	(.036)	(.041)	(.214)	(35.354)
Obs.	1,212	1,212	1,212	1,206
F	3.238	178.6	107.4	35.84
R^2	.002	.339	.341	.362
Adj. R^2	.002	.338	.338	.353

Note: For each model, estimate 'FLE' reflects the marginal effect of the English language treatment on predicted WBS, calculated using the coefficient 'English' and, in Models 3 and 4, the 'English x FLP' coefficient multiplied by the expected value of FLP. The remaining estimates are simply regression coefficients. For all estimates, cluster-robust standard errors are in parentheses.

Table A2: Equivalence Testing Results for Predicted WBS

	Small Effect Size	Medium Effect Size	Large Effect Size
Cohen's *d*	.2	.5	.8
Regression coefficient	.186	.466	.745
FLE *t*-statistic	-2.598	-8.694	-14.790
	(.009)	(< .001)	(< .001)
English x FLP *t*-statistic	1.665	5.780	9.895
	(.048)	(< .001)	(< .001)

Note: Cohen's *d* values are provided directly above the regression coefficients corresponding to their respective Cohen's *d* value. The

equivalence tests here assess whether the observed FLE on predicted WBS is significantly bounded within the range [-*d*, *d*]. *t*-statistics with

corresponding one-sided significance levels in parentheses are calculated using estimates from Model 4 in Table A1. p-values are replaced

with Bonferroni-Holm corrected q-values to correct the family-wise error rate (Holm 1979; Newson 2010).

Appendix D: Main Results by Vignette

Table A3: Main Regressions by Vignette

	(1) Inspector	(2) Negotiator	(3) Whistleblower
FLE	.116	-.038	.098
	(.080)	(.101)	(.078)
English	.978	-.879	1.175
	(.463)	(.603)	(.460)
English x FLP	-.212	.208	-.264
	(.112)	(.149)	(.114)
FLP	-.025	-.040	.114
	(.094)	(.138)	(.092)
FLA	-.030	-.050	.014
	(.061)	(.089)	(.061)
Birth year	.014	-.028	-.073
	(.032)	(.039)	(.037)
Female	-.051	-.117	-.244
	(.088)	(.116)	(.085)
Second-year student	.020	-.042	-.077
	(.089)	(.109)	(.094)
Years of work experience	-.026	.016	-.015
	(.025)	(.029)	(.023)
Employed in public sector	.171	-.004	.283
	(.130)	(.149)	(.110)
Employed in private sector	.145	-.038	.162
	(.123)	(.133)	(.099)
SVO prosocial choices	-.014	-.011	-.017
	(.011)	(.013)	(.011)
SVO competitive choices	.012	-.019	.018
	(.021)	(.054)	(.039)
PSM	-.037	-.060	-.105
	(.049)	(.064)	(.051)
Inspector-negotiator pair	-.162	-.071	
	(.081)	(.101)	
Inspector-whistleblower pair			.030
			(.077)
Constant	-24.566	60.713	149.112
	(63.235)	(78.468)	(73.444)
Obs.	397	412	397
F	2.357	0.927	3.606
R^2	.059	.030	.103
Adj. R^2	.025	-.005	.07

Note: For each model, estimate 'FLE' reflects the marginal effect of the English language treatment on WBS in the vignette sample denoted by the column, calculated using

the coefficient 'English' and the 'English x FLP' coefficient multiplied by the expected value of FLP. The remaining estimates are simply regression coefficients. For all

estimates, cluster-robust standard errors are in parentheses.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table A4: Equivalence Testing Results by Vignette

	Inspector Vignette	Negotiator Vignette	Whistleblower Vignette
Cohen's *d*	.2	.2	.2
Regression coefficient	.214	.214	.214
FLE *t*-statistic	-1.221	2.500	-1.478
	(.446)	(.038)	(.350)
English x FLP *t*-statistic	.012	-.038	-.445
	(1.000)	(1.000)	(1.000)

Note: Cohen's *d* values are provided directly above the regression coefficients corresponding to the Cohen's *d* value for WBS across the entire

sample, regardless of vignette. The equivalence tests here assess whether the observed FLE on WBS is significantly bounded within the range

[-*d, d*]. *t*-statistics with corresponding one-sided significance levels in parentheses are calculated using the respective FLE estimates from

Table A3. *p*-values are replaced with Bonferroni-Holm corrected *q*-values to correct the family-wise error rate across all six tests (Holm 1979;

Newson 2010).

Notes

Note 1: To keep our paper at reasonable length, we refrain from reviewing all mechanisms, focusing only on those relevant to the current study. Other proposed mechanisms include decreased emotional processing under FL usage (Mækelæ & Pfhul 2019) or cultural accommodation to the language in which one is processing information (Akkermans et al. 2010; Pan & Patel 2018).

Note 2: The moral FLEs found in the aforementioned accounting studies are likely confounded because they reflect both the FLE on IFRS interpretation **and** the general moral FLE.

Note 3: This likely has little impact on conclusions from the present FLE literature because the bulk of FLE experiments documented in contemporary meta-analytic reviews (Circi et al. 2021; Stankovic et al. 2022) do not adjust for pre-treatment covariates in their analyses.

Note 4: We originally intended to treat vignette assignment as a proxy for the **shade** of bribery (see, for example, De Waele et al. 2021; Weißmüller & De Waele 2022) and to report treatment effects on the bribery vignettes as reflective of this shade. However, we recognize *post hoc* that we changed the context, monetary stakes, and social consequences of the scenario simultaneously (see Appendix B). This confounds bribery shade treatment effects to such an extent that we decided to not report vignette effects as 'bribery shade' effects. We still report vignette effects regardless, both for transparency and because vignette assignment explains much of the variation in WBS.

Note 5: Recruited participants take part in both this experiment and a second FLE vignette experiment as part of a broader investigation into the determinants of corruption. This second

experiment is not detailed in this study as its vignettes exclusively concern the public sector, and are thus irrelevant to the business ethics literature.

Note 6: This 6.5% figure includes participants excluded for implicit attention check failures; that is, participants who did not answer all required questions in the survey.

Note 7: While we collect data on 33 pre-treatment covariates in service of the broader research project on corruption micro-foundations (see Note 5), we deem only the selected subset to be theoretically relevant for this particular study.

Note 8: Gender identification is voluntary. Three participants who are native Dutch speakers and passed all attention checks refrained from indicating gender, and thus gender data is missing for these individuals. These participants are dropped from the sample in a list-wise fashion in our regression specification that controls for gender.

Note 9: The FLE treatment effect estimator is small enough that a small effect size ($d = 0.2$, a regression coefficient of 0.214) would yield a t-statistic around 4.

Note 10: While Lakens et al. (2018) recommend against this practice, it is not possible to obtain priors for reasonable effect sizes from the extant literature because the WBS measure in this study is novel.

Note 11: While the q-value for the equivalence test of the FLE's effect size is above recently recommended statistical significance thresholds of 0.5% (Benjamin et al. 2018) at a small effect size, this q-value lies below traditional 5% significance thresholds in business studies and the

65

social sciences. We thus follow the recommendations of Benjamin et al. (2018) and classify our FLE equivalence test finding at the small effect size as suggestive.

Note 12: While a monotonic transformation of the vignette effect coefficients for benefit to the briber would match the **signs** of the estimates found in Table 5, **conclusions** would not align because perceived benefit to the briber is significantly higher for the whistleblower vignette than for the inspector vignette ($t > 5$), whereas the difference in WBS between the whistleblower and inspector vignettes is not statistically significant.

Note 13: The risk FLE experiments in Circi et al. (2021) that use task-related incentives include Winskel et al. (2016) and Hayakawa et al. (2019). Keysar et al. (2012) employ both an incentivized experiment (Spanish-English) and an unincentivized experiment (English-Korean).

References

Abbink, K., Freidin, E., Gangadharan, L., & Moro, R. (2018). The effect of social norms on bribe offers. *The Journal of Law, Economics, & Organization* 34(3), 457-474.

Akkermans, D., Harzing, A., & van Witteloostuijn, A. (2010). Cultural accommodation and language priming. *Management International Review* 50(5), 559-583.

Alatas, V., Cameron, V., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Gender, culture, and corruption: Insights from an experimental analysis. *Southern Economic Journal* 75(3), 663-680.

Alempaki, D, Doğan, G., & Yang, Y. (2021). Lying in a foreign language? *Journal of Economic Behavior & Organization* 185, 946-961.

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ* 311, 485.

Argandoña, A. (2003). Private-to-private corruption. *Journal of Business Ethics* 47, 253-267.

Argandoña, A. (2005). Corruption and companies: The use of facilitating payments. *Journal of Business Ethics* 60, 251-264.

Bahoo, S., Alon, I., & Paltrinieri, A. (2020). Corruption in international business: A review and research agenda. *International Business Review* 29, 101660. https://doi.org/10.1016/j.ibusrev.2019.101660.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: Moral frames and social norms. *Experimental Economics* 19, 240-267.

Bartlett, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika* 38(3/4), 337-344.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson,

C., Fehr, E., Fidler, F., Field, A. P., Forster, M., George, E. I., Gonzalez, R., Goodman, S., Green, E., Green, D. P., Greenwald, A. G., Hadfield, J. D., Hedges, L. V., Held, L., Ho, T.-H., Hoijtink, H., Hruschka, D. J., Imai, K., Imbens, G., Ioannidis, J. P. A., Jeon, M., Jones, J. H., Kirchler, M., Laibson, D., List, J., Little, R., Lupia, A., Machery, E., Maxwell, S. E., McCarthy, M., Moore, D. A., Morgan, S. L., Munafó, M., Nakagawa, S., Nyhan, B., Parker, T. H., Pericchi, L., Perugini, M., Rouder, J., Rousseau, J., Savalei, V., Schönbrodt, F. D., Sellke, T., Sinclair, B., Tingley, D., Van Zandt, T., Vazire, S., Watts, D. J., Winship, C., Wolpert, R. L., Xie, Y., Young, C., Zinman, J., & Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour* 2(1), 6–10.

Bereby-Meyer, Y., Hayakawa, S., Shalvi, S., Corey, J. D., Costa, A., & Keysar, B. (2020). Honesty speaks a second language. *Topics in Cognitive Science* 12(2), 632-643.

Białek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology* 85, 103855. https://doi.org/10.1016/j.jesp.2019.103855.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review* 19(5), 547-556.

Bogaert, S., Boone, C., & van Witteloostuijn, A. (2012). Social value orientation and climate strength as moderators of the impact of work group cooperative climate on affective commitment. *Journal of Management Studies* 49(5), 918–944.

Campbell, H., & Gustafson, P. (2018). Conditional equivalence testing: An alternative remedy for publication bias. *PLoS ONE* 13(4). https://doi.org/10.1371/journal.pone.0195145.

Chen, H.-C., Lin, S.-Y., & Chen, I.-H. (2024). Better safe than sorry: The role of anticipated regret and organizational ethical climate in predicting workplace safety behavior. *Chinese Management Studies* 18(1), 48-70.

Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*. https://doi.org/10.1177/00491241221099552.

Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review* 28(4), 1131–41.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

College voor Toetsen en Examens (2023). *Moderne vreemde talen HAVO: Syllabus centraal examen 2025 Arabisch, Duits, Engels, Frans, Russisch, Spaans, Turks*. Technical report, College voor Toetsen en Examens, Utrecht, Netherlands. https://www.examenblad.nl/system/files/2023/syllabi/syllabus_moderne_vreemde_talen_vwo_2025_versie_2.pdf.

Costa, A., Foucart, F., Arnon, I., Aparici, M., & Apesteguia, J. (2014). 'Piensa' twice: On the foreign language Effect in decision making. *Cognition* 130(2), 236–254.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16(3), 297-334.

De Bock, T., & Van Kenhove, P. (2011). Double standards: The role of techniques of neutralization. *Journal of Business Ethics* 99, 283-296.

De Waele, L., Weißmüller K. S., & van Witteloostuijn, A. (2021). Bribery and the role of public service motivation and social value orientation: A multi-site experimental study in Belgium, Germany and the Netherlands. *Frontiers in Psychology* 12, 655964. https://doi.org/10.3389/fpsyg.2021.655964.

Dyen, I., Kruskal, J. B., & Black, P. (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82(5), 1-132.

Edwards, A. (2016). *English in the Netherlands: Functions, forms and attitudes*. John Benjamins, Amsterdam.

Eurobarometer (2012). *Europeans and their languages*. Technical report, European Commission, Brussels, Belgium. https://op.europa.eu/en/publication-detail/-/publication/f551bd64-8615-4781-9be1-c592217dad83.

Fabo, B., Beblavý, M., & Lenaerts, K. (2017). The importance of foreign language skills in the labor markets of Central and Eastern Europe: Assessment based on data from online job portals. *Empirica* 44, 487-508.

Fooks, G., Gilmore, A., Collin, J., Holden, C., & Lee, K. (2013). The limits of corporate social responsibility: Techniques of neutralization, stakeholder management and political CSR. *Journal of Business Ethics* 112, 283-299.

Ford, R. C., & Richardson, W. D. (1994). Ethical decision making: A review of the empirical literature. *Journal of Business Ethics* 13(3), 205-221.

Gardiner, J.A. (2001). Defining corruption. In *Political corruption: Concepts and contexts*. Transaction Publishers, New York, NY.

Gargalianou, V., Muehlfeld, K., Urbig, D., & van Witteloostuijn, A. (2016). Foreign language anxiety in professional contexts: A short scale and evidence of personality and gender differences. *Schmalenbach Business Review* 17, 195-223.

Gärling, T., Fujii, S., Gärling, A., & Jakobsson, C. (2003). Moderating effects of social value orientation on determinants of proenvironmental behavior intention. *Journal of Environmental Psychology* 23(1), 1–9.

Gatti, R., Paternostro, S., & Rigolini, J. (2003). Individual attitudes toward corruption: Do social effects matter? World Bank Policy Research Working Paper 3122. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-3122.

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

Ginsburgh, V., & Weber, S. (2016). Linguistic distances and ethnolinguistic fractionalization and disenfranchisement indices. In Ginsburgh, V., & Weber, S. (eds.), *The Palgrave handbook of economics and language*, (pp. 137–173). Palgrave Macmillan, London, England.

Goel, R. K., Budak, J., & Rajh, E. (2015). Private sector bribery and effectiveness of anti-corruption policies. *Applied Economics Letters* 22, 759-766.

Gopinath, C. (2008). Recognizing and justifying private corruption. *Journal of Business Ethics* 82, 747-754.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics* 124, 695-726.

Gross, J., Leib, M., Offerman, T., & Shalvi, S. (2018). Ethical free riding: When honest people find dishonest partners. *Psychological Science* 29(12), 1956-1968.

Hagen, S., Foreman-Peck, J., Davila-Philippon, S., Nordgren, B., & Hagen, S. (2006). *ELAN: Effects on the European economy of shortages of foreign language skills in enterprise*. Technical report, The National Centre for Languages (CILT), Reading, United Kingdom. https://ec.europa.eu/assets/eac/languages/policy/strategic-framework/documents/elan_en.pdf.

Hahm, S., & Gazzola, M. (2022). The value of foreign language skills in the German labor market. *Labour Economics* 76, 102150. https://doi.org/10.1016/j.labeco.2022.102150.

Hayakawa, S., Lau, B. K. Y., Holtzmann, S., Costa, A., & Keysar, B. (2019). On the reliability of the foreign language effect on risk-taking. *Quarterly Journal of Experimental Psychology* 72(1), 29-40.

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science* 28(10), 1387-1397.

Heidenheimer, A. J. (1970). *Political corruption: Readings in comparative analysis*.heidenhei

    Routledge, New York, NY.

Hensel, P. G. (2021). Reproducibility and replicability crisis: How management compares to

    psychology and economics – A systematic review of literature. *European Management*

    Journal 39(5), 577-594.

Herics, O., Obermayr, T., Puricella, P., Grassi, E., Fara, G., Hai, D., Hristov, S., Jiménez, C.,

    Latopoulou, C., & Ploumaki, M. (2018). *Public private partnerships in the EU:*

    Widespread shortcomings and limited benefits. Technical report, European Court of

    Auditors, Luxembourg, Luxembourg. https://op.europa.eu/webpub/eca/special-

    reports/ppp-9-2018/en/.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal*

    of Statistics 6(2), 65–70.

Huang, J.-Y., & Chen, C.-A. (2023). Relationship between criminal intention and criminal

    behavior. *Journal of Police and Criminal Psychology* 38, 555-566.

Jaakson, K., Johannsen, L., Pedersen, K. H., Vadi, M., Ashyrov, G., Reino, A., & Sööt, M.-L.

    (2019). The role of costs, benefits, and moral judgments in private-to-private

    corruption. *Crime, Law and Social Change* 71, 83-106.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, NY.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika* 35(4), 401-415.

Kankaanranta, A., & Lu, W. (2013). The evolution of English as the business lingua franca:

    Signs of convergence in Chinese and Finnish professional communication. *Journal of*

    Business and Technical Communication 27(3), 288-307.

Kaufmann, D., Montorial-Garriga, J., & Recanatini, F. (2008). How does bribery affect public

    service delivery? Micro-evidence from service users and public officials in Peru. World

Bank Policy Research Working Paper 4492. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-4492.

Keig, D. L., Brouthers, L. E., & Marshall, V. B. (2015). Formal and informal corruption environments and multinational enterprise social irresponsibility. *Journal of Management Studies* 52(1), 89-116.

Kepes, S., Banks, G. C., & Oh, I. S. (2014). Avoiding bias in publication bias research: The value of "null" findings. *Journal of Business and Psychology* 29, 183-203.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science* 23(6), 661-668.

Kim, S., Vandenabeele, W., Wright, B. E., Andersen, L.B., Cerase, F. P., Christensen, R. K., Desmarais, C., Koumenta, M., Leisink, P., Liu, B., Palidauskaite, J., Pedersen, L. H., Perry, J. L., Ritz, A., Taylor, J., & De Vivo, P. (2013). Investigating the structure and meaning of public service motivation across populations: Developing an international instrument and addressing issues of measurement invariance. *Journal of Public Administration Research and Theory* 23(1), 79-102.

Köbis, N. C., van Prooijen, J.-W., Righetti, F., & Van Lange, P. A. M. (2015). "Who doesn't?" – The impact of descriptive norms on corruption. *PLoS ONE* 10(6), e0131830. https://doi.org/10.1371/journal.pone.0131830.

Lakens, D., Scheel, A.M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science* 1(2), 259-269.

Lee, S.-H., & Hong, S. J. (2012). Corruption and subsidiary profitability: US MNC subsidiaries in the Asia Pacific region. *Asia Pacific Journal of Management* 29, 949-964.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal* 92(368), 805-824.

Mækelæ, M. J., & Pfuhl, G. (2019). Deliberate reasoning is not affected by language. *PLoS ONE* 14(3), e0211428. https://doi.org/10.1371/journal.pone.0211428.

Marschan-Piekkari, R., Welch, D., & Welch, L. (1999). Adopting a common corporate language: IHRM implications. *International Journal of Human Resource Management* 10(3), 377-390.

McDonald, G. (2000). Cross-cultural methodological issues in ethical research. *Journal of Business Ethics* 27(1/2), 89-104.

Melitz, J. (2016). English as a global language. In Ginsburgh, V., & Weber, S. (eds.), *The Palgrave handbook of economics and language*, (pp. 583-615). Palgrave Macmillan.

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science* 62(3), 760-765.

Newson, R. B. (2010). Frequentist q-values for multiple-test procedures. *The Stata Journal* 10(4), 568-584.

Pan, P., & Patel, C. (2018). The influence of native versus foreign language on Chinese subjects' aggressive financial reporting judgments. *Journal of Business Ethics* 150(3), 863-878.

Park, J., & Lee, K.-H. (2020). Organizational politics, work attitudes and performance: The moderating role of age and public service motivation (PSM). *International Review of Public Administration* 25(2), 85-105.

Paternoster, R., Bachman, R., Bushway, S., Kerrison, E., & O'Connell, D. (2015). Human agency and explanations of criminal desistance: Arguments for a rational choice theory. *Journal of Developmental and Life-Course Criminology* 1, 209-235.

Paternoster, R., & Simpson, S. (1996). Sanction threats and appeals to morality: Testing a rational choice model of corporate crime. *Law & Society Review* 30(3), 549-584.

Peterson, D., Rhoads, A., & Vaught, B. C. (2001). Ethical beliefs of business professionals: A study of gender, age and external factors. *Journal of Business Ethics* 31, 225-232.

Pletti, C., Lotto, L., Tasso, A., & Sarlo, M. (2016). Will I regret it? Anticipated negative emotions modulate choices in moral dilemmas. *Frontiers in Psychology* 7, 01918. https://doi.org/10.3389/fpsyg.2016.01918.

Potipiroon, W., & Faerman, S. (2016). What difference do ethical leaders make? Exploring the mediating role of interpersonal justice and the moderating role of public service motivation. *International Public Management Journal* 19(2), 171-207.

Powpaka, S. (2002). Factors affecting managers' decision to bribe: An empirical investigation. *Journal of Business Ethics* 40, 227-246.

Ramdani, D., & van Witteloostuijn, A. (2012). The shareholder-manager relationship and its impact on the likelihood of firm bribery. *Journal of Business Ethics* 108, 495-507.

Ramdani, D., & van Witteloostuijn, A. (2014). Bribery. In Arrington, C. B., Blowers, A. N., Brennan, P. K., Brewster, M. P., Bumgarner, J. B., Cencich, J. R., Cordner, A., Dodge, M., Joseph, J., Kurlychek, M. C., McConnell, E. H., Nasheri, H., Roth, M. P., Schneider, J. L. (eds.), *The Encyclopedia of Criminology and CriminalJjustice*. Blackwell. https://doi.org/10.1002/9781118517383.wbeccj066.

Robertson, D. C., & Nichols, P. M. (2017). Introduction and overview: Bribery and the study of decision making. In Robertson, D. C., & Nichols, P. M. (eds.), *Thinking about bribery: Neuroscience, moral cognition and the psychology of bribery*, (pp. 1-32). Cambridge University Press.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling*. Routledge, London.

Senci, C. M., Hasrun, H., Moro, R., & Freidin, E. (2019). The influence of prescriptive norms and negative externalities on bribery decisions in the lab. *Rationality and Society* 31(3), 287-312.

Stankovic, M., Biedermann, B., & Hamamura, T. (2022). Not all bilinguals are the same: A meta-analysis of the moral foreign language effect. *Brain and Language* 227, 105082. https://doi.org/10.1016/j.bandl.2022.105082.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23(5), 645-665.

Sugahara, S., Tsunogaya, N., & Kim, J.-H. (2023). Interaction effect of foreign language and obedience pressure on ethical judgment in accounting: Evidence from Japan. *Journal of International Accounting Research* 22(1), 83-98.

Swamy, A., Knack, S., Lee, Y., & Azfar, O. (2001). Gender and corruption. *Journal of Development Economics* 64(1), 25-55.

Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review* 22(6), 664-670.

Urbig, D., Terjesen, S., Procher, V., Muehlfeld, K., & van Witteloostuijn, A. (2016). Come on and take a free ride: Contributing to public goods in native and foreign language settings. *Academy of Management Learning & Education* 15(2), 268-286.

van Dijk, E., & De Cremer, D. (2006). Self-benefiting in the allocation of scarce resources: Leader-follower effects and the moderating effect of social value orientations. *Personality and Social Psychology Bulletin* 32(10), 1352-1361.

van Hugten, J., & van Witteloostuijn, A. (2018). The foreign language effect on the self-serving bias: A field experiment in the high school classroom. *PLoS ONE* 13(2), e0192143. https://doi.org/10.1371/journal.pone.0192143.

van Prooijen, J.-W., De Cremer, D., van Beest, I., Ståhl, S., van Dijke, M., & Van Lange, P. A. M. (2008). The egocentric nature of procedural justice: Social value orientation as moderator of reactions to decision-making procedures. *Journal of Experimental Social Psychology* 44(5), 1303-1315.

van Witteloostuijn, A. (2016). What happened to Popperian falsification? Publishing neutral and negative findings: Moving away from biased publication practices. *Cross Cultural & Strategic Management* 23(3), 481-508.

Warner, C. H., Fortin, M., & Melkonian, F. (2022). When are we more ethical? A review and categorization of the factors influencing dual-process ethical decision-making. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-022-05281-0.

Weißmüller, K. S., & De Waele, L. (2022). Would you bribe your lecturer? A quasi-experimental study on burnout and bribery in higher education. *Research in Higher Education* 63(5), 768-796.

Weißmüller, K. S., Bouwman, R., & Vogel, R. (2023). Satisficing or maximizing in public-private partnerships? A laboratory experiment on strategic bargaining. *Public Management Review* 25(7), 1282-1308.

Weißmüller, K. S., & Zuber, A. (2023). Understanding the micro-foundations of administrative corruption in the public sector: Findings from a systematic literature review. *Public Administration Review* 83, 1704-1726.

Werner, S. B. (2001). The development of political corruption in Israel. In *Political corruption: Concepts and contexts*. Transaction Publishers, New York, NY.

Winskel, H., Ratitamkul, Brambley, T., Nagarachinda, T., & Tiencharoen, S. (2016). Decision-making and the framing effect in a foreign and native language. *Journal of Cognitive Psychology* 28(4), 427-436.

Youyou, W., Yang, Y., & Uzzi, B. (2023). A discipline-wide investigation of the replicability of psychology papers over the past two decades. *PNAS* 120(6), e2208863120. https://doi.org/10.1073/pnas.2208863120.