Bilingualism: Language and Cognition

cambridge.org/bil

Effects of the French grammatical gender system on bilingual adults' perception of objects

Zhuohan Chen and Faidra Faitaki

Department of Education, University of Oxford, Oxford, UK

Research Article

Cite this article: Chen, Z., & Faitaki, F. (2025). Effects of the French grammatical gender system on bilingual adults' perception of objects. *Bilingualism: Language and Cognition*, **28**, 401–414. https://doi.org/10.1017/S1366728924000464

Received: 16 January 2024
Revised: 13 May 2024
Accepted: 26 May 2024
First published online: 3 October 2024

Keywords:
linguistic relativity; psycholinguistics; French grammatical gender system; language and perception; bilingualism and cognition

Corresponding author:
Zhuohan Chen;
Email: zhuohan.chen@education.ox.ac.uk

Abstract

This study extends the line of linguistic relativity research by assessing the effect of the French grammatical gender system on French speakers' and learners' perception of objects. Four groups of 140 adults (English monolinguals, French monolinguals, English–French bilinguals and French–English bilinguals; $N = 35$ each) rated 32 selected objects' gender by assigning them a masculine/feminine voice on a slider. We also assessed the participants' second-language (L2) proficiency. Multilevel modelling results revealed that French monolinguals and English–French bilinguals rated objects' gender in line with the French grammatical gender system. The effect of French on perception was not reduced by acquiring English, as French–English bilinguals performed on par with French monolinguals. Moreover, the effect was independent of L2 proficiency. These findings suggest that learning a gendered L2 affects the perception of objects – thus supporting the linguistic relativity hypothesis.

1. Introduction

The past few decades have seen a remarkable resurgence of interest in linguistic relativity (Sapir, 1956; Whorf, 1956), the hypothesis that language influences thoughts (Athanasopoulos & Casaponsa, 2020). Substantial debate surrounds linguistic relativity; specifically, to what extent people's thoughts can be biased by language (Casasanto, 2008; Pinker, 1994). A growing body of research exploring linguistic relativity has been focusing on the effect of grammatical gender systems on perception (Bassetti & Nicoladis, 2016; Lambelet, 2016; Samuel et al., 2019; Sato & Athanasopoulos, 2018). A grammatical gender system assigns genders to inanimate nouns. The link between the assigned gender and the referent's properties is arbitrary, rendering the gender system challenging to learn (Bassetti & Nicoladis, 2016; Shimanskaya & Slabakova, 2019). For example, the French word *table* ('la table') is feminine but has no prominent association with femininity. Therefore, studies have examined whether speakers of the gendered first language (L1) would think of the objects' gender in line with the grammatical gender system (e.g., Bender et al., 2011).

Earlier studies show there could be a potential effect of a language's grammatical gender system on its monolingual speakers' object perception, as well as on the object perception of the second-language (L2) learners with a genderless L1 (e.g., Lambelet, 2016). For example, a French monolingual speaker might perceive the object *table* as feminine, i.e., in line with its grammatical gender in French. Similarly, an English learner of French might also think of table as feminine because of having acquired French. The impact of a grammatical gender system on speakers of a gendered L1 seems to remain unaffected by the acquisition of a genderless L2 like English (e.g., Bassetti & Nicoladis, 2016). However, learning other grammatical gender systems might change this impact. For example, a French learner of German might think of the object *table* differently because *table* is masculine in German.

Nevertheless, methodological issues exist in relevant studies. It is difficult to know whether the effect of the gender system on perception remains when the bilinguals recruited can use the L2 effortlessly. Indeed, it is also uncertain whether a higher level of L2 proficiency would increase, decrease or have no effect on the impact of the L1 grammatical gender system. Moreover, it is inconclusive whether the L1 grammatical gender system still impacts perception after confounding variables like gender stereotypes associated with objects are adequately controlled within the research design. These methodological limitations render the results of earlier studies potentially limited regarding the strength of inference.

1.1 Language and thoughts: linguistic relativity

There seems to be empirical evidence for linguistic relativity from cognitive linguistic research (Flecken et al., 2015; Kuo & Sera, 2009; Lakoff & Johnson, 1980; Pandey, 2017; Pavlenko, 2003;

This article has earned badges for transparent research practices: Open Materials. For details see the Data Availability Statement.

Perlovsky, 2009). The most frequently cited evidence is on colour terminology (e.g., Athanasopoulos, 2009). For example, Brown and Lenneberg (1954) reported that Zuñi speakers had greater difficulties in distinguishing nuances in colours between blue and green than did English speakers. These difficulties were attributed to the lack of classification for blue and green colours in Zuñi. Linguistic relativity also applies to foreign language learning. The increased emotional distance between the foreign language and the L1 might affect the decision-making and reasoning processes in the foreign language; this is known as the 'foreign language effect' (Chen, 2020; Keysar et al., 2012). For example, reading in a foreign language might decrease one's belief in conspiracy theories because the greater emotional distance might lead to more rational thinking (Chen, 2020). However, linguistic relativity has received much criticism (e.g., Bickel, 2000). There is a consensus that language may affect thinking, but different linguists differ in the degree they believe it does (Ahearn, 2017; Samuel et al., 2019).

As scholars explore linguistic relativity, there has been a lack of consistency in their definitions of the outcome of the hypothesis, i.e., 'thoughts'. Such inconsistency in the definition of 'thoughts' could be problematic for the relativity debate, especially since the various definitions differ on how measurable the outcome is. Many terms have been used in the literature to denote thoughts, including 'ways of thinking', 'cognition', 'worldviews', 'cognitive processes' and 'perception' (e.g., Athanasopoulos, 2009; Casasanto, 2008; Flecken et al., 2015). We argue that some of these constructs are interrelated but are not identical (Firestone & Scholl, 2016) and thus cannot be used interchangeably. Among them, perception appears to be the easiest to measure: it is easier to see how participants perceive something, as opposed to what they thought or what their cognitive states were. Indeed, 'perception' can be the basis of 'thoughts' or 'cognition' which involves more latent nuances. Therefore, in this study, *perception* was chosen as the key term to describe the potential conceptual changes resulting from the effect of language.

1.2 Effects of grammatical gender on object perception

1.2.1 Grammatical gender systems

A grammatical gender system provides a good case for linguistic relativity research because it can investigate the 'pure' linguistic effects on perception 'without the potential confounding effects of non-linguistic cognition' (Bassetti, 2007, p. 254). Many studies have supported the effect of a grammatical gender system on children's and adults' perceptions of object gender (Bassetti, 2007; Clarke et al., 1981; Flaherty, 2001). Nevertheless, the results are often inconclusive because of other variables, as shown by a literature review conducted by Bassetti and Nicoladis (2016), in which the effect of the gender system on perception was found to rely on the pair of languages tested, the language proficiency of the participants and the choice of tasks.

Another systematic review conducted by Samuel et al. (2019) also claimed that the effect of a gender system on perception was context- and task-dependent. Samuel et al.'s (2019) review included 43 empirical studies on the effect of grammatical gender systems on perception from 1990 to 2018, with 5,895 participants (monolinguals, bilinguals and multilinguals). The review reported mixed findings, with 38% of samples indicating support for the effect, 28% showing mixed support and the remaining 34% showing no support. Of all the tasks used in the included studies, voice attribution tasks (assigning a female or male voice to objects) and sex assignment tasks tended to yield the most supportive evidence. In comparison, the properties judgement task (describing an object's properties) yielded the most disapproving evidence against the idea that grammatical gender influences perception.

Samuel et al. (2019) also concluded that the effect of a gender system is influenced by potential confounding variables; for example, how participants make sex-related judgements. In sex assignment and voice attribution tasks, and possibly in tasks involving similarity, association and object–name memory, if participants utilised the grammatical gender system as a strategy for determining the gender of an object, their actions could naturally generate supporting evidence for the relativity hypothesis. Notably, in one study participants admitted to using grammatical gender to guide their responses (Almutrafi, 2015). This suggests that participants relied on a conscious metalinguistic strategy, which in turn undermines an interpretation of the results. It is also possible that participants assign genders to objects based on the grammatical gender of the objects to provide responses they believe meet the aim of the study. Consequently, if studies fail to incorporate mechanisms to control for this tendency, the conclusions obtained would be limited.

1.2.2 Monolinguals' perceptions affected by a gendered language

According to linguistic relativity, speakers of a language with a grammatical gender system would perceive the objects' gender consistently with their L1. Such an effect was found in Spanish (Konishi, 1993), Arabic (Clarke et al., 1981), French (Sera et al., 2002), Italian (Vigliocco et al., 2005) and Lithuanian (Vernich et al., 2017), among other languages. However, the effect was less transparent in German, which has three grammatical genders (masculine, feminine and neuter). It is generally believed that the gender system in German tends to have a less potent effect on object perception, because of the lexical-semantic complexity of its case system (Sera et al., 2002; Vigliocco et al., 2005; Zubin & Köpcke, 1984). Using a neuter gender might have a weaker effect than languages that use a dichotomous gender system (Bassetti, 2007).

For example, Flaherty (2001) conducted three consecutive experiments to explore how the grammatical gender system of Spanish affected the speakers' perception of objects. Flaherty's (2001) experiment investigated the effect of language on gender attribution, involving 144 native English-speaking and 144 native Spanish-speaking adults and children. The study used a robust design to examine the role of age: 48 adults, 48 children aged five to seven and 48 children aged eight to ten, with an equal gender distribution among participants. Participants were asked to assign a typical male or female name to black-and-white cartoons of 35 objects. Results revealed significant consistency between the grammatical gender in Spanish and the participants' choice of gender in the Spanish 8- to 10-year-old children and adults, but not in 5- to 7-year-olds. Spanish children aged five to seven tended to assign gender to objects in line with their own biological gender or with famous characters in children's literature. By contrast, English participants marked the objects' gender according to their characteristics or properties. The results indicated that language affected the perception and that there was an age boundary (at around eight), after which the language's effect began to play a role.

1.2.3 Monolinguals versus bilinguals: from genderless to gendered

Learning a grammatical gender system in a new language tends to affect the perception of native speakers of a genderless language (Bassetti & Nicoladis, 2016). For example, a longitudinal

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

quantitative study by Kurinski and Sera (2011) showed that native English speakers' gender attribution to inanimate objects was affected after learning the Spanish grammatical gender system over 1 academic year. However, although native monolingual Spanish speakers exhibited a strong tendency to assign gender to objects in accordance with the Spanish grammatical gender, English learners of Spanish also showed a similar tendency, albeit not as strongly pronounced; and the effect did not increase in tandem with the learners' proficiency in Spanish (Kurinski & Sera, 2011; also see Wasserman & Weseley, 2009). This finding suggests that the acquisition of a gendered language might already affect learners' perception of a genderless L1 from the early stages of L2 acquisition, but the effect plateaus afterwards.

1.2.4 Monolinguals versus bilinguals: from gendered to genderless

The effect of the L1 gender system on bilinguals with a gendered L1 and a genderless L2 appears to remain unaffected (Bassetti & Nicoladis, 2016). For example, Boroditsky and Schmidt (2000) found that German–English and Spanish–English bilinguals still perceive object gender based on their L1s. However, this study did not report on the English proficiency of these bilinguals. This made the interpretation of the results unclear, as the level of proficiency at which the effect plateaus, and whether the effect changes because of having high L2 ability were unknown. The same 'unwavering' impact of learning a genderless L2 as a learner with a gendered L1 was also found in French–English bilingual adults and children, who aligned their gender perception of objects with the French grammatical gender system (Forbes et al., 2008; Nicoladis & Foursha-Stevenson, 2012; Sato et al., 2013).

Whether the effect of the gender system decreases or increases with L2 proficiency is inconclusive. Sato et al. (2013) reported that French–English bilingual adults' stereotypical gender attitudes were impacted less by the French grammatical gender system as their English proficiency increased. However, Kurinski and Sera (2011) found that an increase in Spanish ability did not affect the perception of English learners of Spanish. Sato et al.'s (2013) findings might be more reliable than those of Kurinski and Sera (2011) because Sato et al. (2013) adopted a standardised c-test to establish L2 proficiency and divided their participants into two groups (advanced and intermediate) more objectively. By contrast, Kurinski and Sera (2011) did not test the learners' Spanish levels; they assumed they were advanced because they were students in the Department of Spanish. The lack of proper L2 proficiency testing has been a common limitation in earlier studies, which have relied on participants' self-evaluation. This project further explores proficiency's role in the effects of the grammatical gender system.

1.2.5 Monolinguals versus bilinguals: having more than one 'gender'

The question of whether the L1 would be affected by the L2 has also been the focus of research regarding having two gendered languages. Bassetti (2007) investigated whether Italian–German bilinguals and Italian monolinguals had different concepts of the same object when the object had opposite genders in German and Italian. Twenty-one Italian–German bilingual and 21 Italian monolingual 9-year-old children (control) participated in an online voice attribution task in which they assigned a male or female voice to 12 concrete objects (choosing from two voice files). All the children were native Italian speakers living in Italy. Results showed that grammatical gender significantly affected children's perception of objects, as indicated by monolingual children's preference to assign a gender consistently with the Italian grammatical gender system.

Similarly, Lambelet (2016) discussed the difficulties bilinguals encounter when learning a second grammatical gender system. A voice attribution task was distributed to 282 adults French L2 learners with 21 different L1 backgrounds, including languages with binary or tertiary grammatical gender systems and languages without it. Participants chose a female or male voice for ten inanimate objects, followed by a good distractor to decide whether the voice belonged to an old or young person. A significant effect of the L1 grammatical gender on the participants' voice attribution task performance was found. Moreover, learning an L2 grammatical gender was found to weaken the impact of L1 grammatical gender: participants with a higher proficiency in the French grammatical gender system were less inclined to assign voices based on the grammatical gender of the object in their L1. The study included a second task, French grammatical gender identification, in which participants were required to identify the French gender of objects used in the experiment. The task found that the conjectured French grammatical gender played a role in deciding gender. For example, *beer* is neutral in German ('das Bier'), but, if a German participant incorrectly thought *beer* in French was masculine (in fact, it is feminine), s/he would assign a masculine gender, indicating the effect of conjectured L2 grammatical gender system. However, these findings should be interpreted with caution because no French proficiency tests were implemented to assess the participants' French level. Also, objects did not involve reasonable control of items with a stereotypical gender, which jeopardises the interpretation of the effect of the L2 gender system. Furthermore, the binary choice between a masculine or feminine voice may present challenges. Participants might find it difficult to categorise the gender of an inanimate object as strictly 'masculine' or 'feminine'. A more effective approach could be to give participants greater flexibility in their choices, as they may lean towards decisions falling somewhere in between.

1.3 Research gap and rationale for the study

In summary, a language's (binary) grammatical gender system appears to affect perception among monolingual speakers and learners with a genderless L1, but its effect depends on contexts and tasks. The L1 tends to remain unaffected by learning a genderless L2 but might be weakened by acquiring another gendered L2. The results of relevant studies are often inconclusive and tentative due to methodological problems, including the lack of L2 proficiency testing, the use of gender stereotypes and the lack of control items for objects' stereotypical gender tendencies. Therefore, this study aims to explore the effect of the French grammatical gender system with improved methods.

1.4 Research questions

The aims of this study are twofold. Firstly, it investigates if (and if so, how) the grammatical gender system in French affects French speakers' perception of objects (e.g., is the *table* perceived as masculine or feminine?) compared to English monolingual speakers. Secondly, it explores if (and if so, how) learning English, an L2 that has/does not have a grammatical gender system, affects bilingual adults' perception of objects compared with monolingual speakers.

The research questions (RQ) are as follows:

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

(1) To what extent do English and French monolingual speakers perceive objects differently (assigning voice; masculine or feminine)?
(2) (L1 genderless vs L2 gender) To what extent do English–French bilingual[1] speakers differ from English monolingual speakers in perceiving objects after learning French?
(3) (L1 gender vs L2 genderless) To what extent do French–English bilingual speakers differ from French monolingual speakers in perceiving objects after learning English?

1.5 Predictions

This study involved two within-participants variables, the chosen objects' French gender (male or female) and condition (experimental items and control items), and one between-participants variable (language group). For RQ1 and RQ2, a main effect of group was expected; English and French monolingual speakers, as well as English monolinguals and English-dominant bilinguals, were expected to differ in their ratings of objects' gender. An interaction effect between group and condition for both RQ1 and RQ2 was also expected: French monolinguals and English-dominant bilinguals would rate control items (measured more stereotypically gendered concepts) based on their characteristics or properties, and experimental items based on the grammatical gender system in French. By contrast, the English monolingual group was predicted to rate all items based on their characteristics or properties. Moreover, an interaction effect between group and gender was expected, suggesting that French monolinguals and English-dominant bilinguals would assign male objects a more masculine voice and female objects a more feminine voice than English monolinguals.

No main effect of the group was expected for RQ3. As shown in earlier literature, learning a genderless language did not appear to influence the impact of the L1 grammatical gender system (e.g., Bassetti & Nicoladis, 2016). Thus, the French–English bilingual group might not differ from the French monolingual group. No interaction effect between group and condition, nor between group and gender was predicted.

2. Methodology

2.1 Participants

A total of 140 participants (35 per group) were recruited through Prolific (a participant recruitment website using pre-screening criteria) and categorised into four groups (English monolinguals, French monolinguals, English-dominant bilinguals and French-dominant bilinguals). The power analysis using the G*Power 3 program (Faul et al., 2007) indicated that 134 participants (33 per group) would be required to achieve 80% power at the $\alpha = .05$ level ($[1 - \beta] = .80$) for main effects.[2] Thus, the study can be deemed as having sufficient power. Included bilingual participants were all manually screened as having L2 proficiencies at least higher than B2[3] and having English/French as L2 (tested via vocabulary knowledge, discussed later).

Participants' gender distribution was relatively equal (51% female), thus mitigating the effect of participants' own gender on their perception (Flaherty, 2001). Among participants, 14% ($N = 10$) were simultaneous bilinguals (one French-dominant bilingual and nine English-dominant bilinguals), whereas 39% ($N = 28$) began learning their L2 between the ages of four and nine, and 47% ($N = 32$) started at age ten or older. The majority of bilingual individuals ($N = 41$) reported having 0–30% daily exposure to their L2, with 27% ($N = 19$) having 30–70% exposure and 14% ($N = 10$) having more than 70% exposure. Additionally, 36% ($N = 25$) of bilingual participants held qualifications (formal recognition of their language ability like language proficiency tests or academic degrees) in their L2. However, there was a discrepancy between English-dominant bilinguals' French level and French-dominant bilinguals' English level as a whole: most of the French participants (69%, $N = 21$) were advanced/proficient in English (C1 and C2), whereas most of the English participants (91%, $N = 32$) had only an upper and low intermediate levels (B1 and B2).

2.2 Materials

2.2.1 Background information questionnaire

The experiment was built on Gorilla Experiment Builder. Participants first completed a background information questionnaire, including the participants' language backgrounds, language exposure, age of onset for L2, self-perceived L2 proficiency, language qualifications, gender, education level and other language learning experiences likely to be moderating variables in this study.

2.2.2 Vocabulary tests

The second part of the experiment included vocabulary tests measuring the bilinguals' proficiency in English or French (Miralpeix & Muñoz, 2018). English-dominant bilinguals completed the French vocabulary test in English, whereas French-dominant bilinguals completed the English tests in French. The vocabulary test also assisted in the manual screening of bilingual participants: a participant needed to attain a vocabulary test score of at least 60% (B1) to be deemed bilingual instead of monolingual.

The vocabulary test used for English was the Lexical Test for Advanced Learners of English (LexTALE), a visual lexical decision task, proven to be a good predictor of one's vocabulary knowledge or even general proficiency in English (Lemhöfer & Broersma, 2012). The French version of LexTALE built by Brysbaert (2013) (LexTALE_Fr) was used (for word lists, refer to Appendix S1). The English LexTALE was only given to French-dominant bilingual speakers, and vice versa. Words were presented individually, and the order was randomised for each participant. Both the English and the French versions involve the same yes/no decision task, asking participants to decide if the word is an actual word. All 70 bilinguals' L2 proficiency was scored manually. The relationship between LexTALE scores and L2 proficiency in this study was based on Lemhöfer and Broersma's (2012) work: a score of 80–100% corresponds to upper and lower advanced or proficient users (C1 and C2). Achieving a score between 60% and 80% places individuals at the upper intermediate level (B2). Those with scores below 59% fall into the lower intermediate and below categories (B1 and lower).

2.2.3 Voice decision task

In the third part of the experiment, participants completed a voice decision task in which they decided on the voice for objects (Lambelet, 2016). For example, 'if the *airplane* could speak, what kind of voice would it have?'. In the French version, instead of *le avion* ('the airplane'), *l'object* ('the object') was used to avoid the hint of the gender pronoun. As shown in Figure 1, participants attributed voices using a slider, with the far-left representing

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

low-pitch (or 'masculine' sound) and the far-right representing high-pitch (or 'feminine' sound). The slider gave rise to continuous data, as the far-left corresponded to 1% and the far-right to 100%. Thus, the point where the participants placed the slider represented a percentage. An object was considered to have a masculine voice if a participant assigned it a score of up to 50%, and the object was considered to have a feminine voice if the participant gave it a score of over 51%. The slider also allowed more space for participants to choose from dichotomous masculine or feminine, thus potentially making them feel more comfortable.

2.2.4 Piloting, stimuli and procedure

The experiment link was compiled and generated on the Gorilla Experiment Builder. Before the primary study, piloting was done with 30 English and 30 French monolingual speakers who were not included in the subsequent experiment. The piloting aimed to ensure internal validity: the ratings participants needed to give for objects were based on personal evaluation, not on the stereotypical association with the object's characteristics or properties (e.g., *necklace* is always feminine, whereas *hammer* is usually masculine, Bassetti, 2007). Thus, piloting allowed the study to determine the objects with an inherent 'gender' attribution (which would correspond to control items in the experiment) and those without (experimental items).

Based on the pilot, the final objects (see Table S1 for object list) included in the experiment were 32 common concrete objects (16 feminine and 16 masculine words in French), some of which were based on Flaherty's (2001) work. All words had high frequency ($M = 6{,}791$) in the British National Corpus, and the mean length was 1.38. Ten control items were chosen (English and French monolinguals rated them ≤40 or≥60 in the piloting), with five in each gender. The remainder, 11 feminine and 11 masculine objects, were experimental items. The ten experimental items rated significantly differently by participants during the pilot were chosen to be the target of open-ended questions (in which participants explained their choices in a text box).

All objects were illustrated using black-and-white simple line drawings drawn by the researcher, controlling for complexity (see Figure S1). The black-and-white format was chosen to eliminate the gender connotations of colour (Flaherty, 2001). No audio files were used in this experiment. This decision was motivated by piloting with audio files showing participants were disturbed by the voice in the file. Ten open-ended questions (serving as attention checks) were presented in the voice attribution task following the ten objects rated significantly differently in the piloting. The question asked, 'why do you think the voice you just rated will be like this?'. Participants had to type their answers in a box.

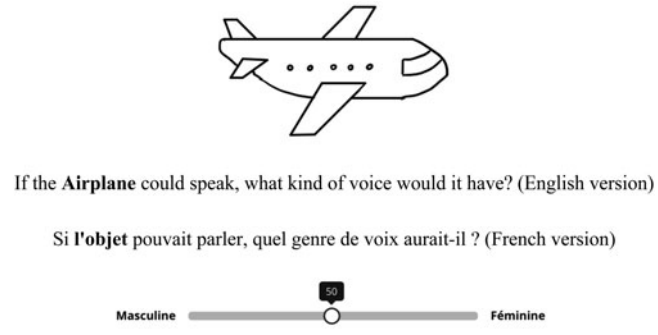

Figure 1. Example of voice attribution task.

English monolinguals and English-dominant bilinguals completed the experiment in English, whereas French monolinguals and French-dominant bilinguals did so in French. Back-translation ensured the highest translation fidelity (Thompson & Dooley, 2019). Consent was obtained at the beginning of the experiment. All data collected were non-identifiable. The consent form stated explicitly that 'this study generally categorises masculine and feminine voices in a continuum for purely academic purposes'. This was to reduce the risk of disagreement from participants in simply categorising voice as having low-pitch vowel sound (generally perceived as the quality of masculine) and high-pitch vowel sound (usually perceived as the quality of feminine).

3. Results

The outcome measure was the 140 participants' ratings for 32 objects. There were categorical predictors regarding the objects were Condition (experimental and control) and ObjectGender (French female or male). There was also one continuous predictor on the individuals LanguageGroup. Each participant's ten open-ended questions (1,400 answers) were coded by the first author according to three levels: 'grammatical gender', 'characteristics and feelings' and 'other'. An independent coder rated all the answers again to reduce the researcher bias, with Cronbach's $\alpha = .87$, which shows high inter-coder reliability (Field, 2018).

3.1 Overall effects of independent variables

Descriptive statistics are shown in Table 1. Multilevel modelling (MLM) was adopted (Hoffman & Rovine, 2007). In this study, two types of random effects (for *objects* and *participants*) were explored. The fixed effects were *LanguageGroup*, *ObjectGender* and *Condition*. The MLM was carried out using R 4.0.4 with the lme4 package (Bates et al., 2015).

Six models were compiled linearly. Models 1, 2 and 3 aimed to determine the fixed main effects of LanguageGroup, Condition and ObjectGender, while controlling the random effects for objects and participants. The next three models aimed to identify the fixed interaction effects of LanguageGroup × ObjectGender, LanguageGroup × Condition and LanguageGroup × Condition × ObjectGender, respectively. The codes can be found in Appendix S2.

Table 2 displays the results for main and interaction effects. In model 1, the main effect of LanguageGroup only appeared when comparing the French monolingual group to the English monolingual group ($M_{change} = 4.356$, $p < .05$). No main effect of Condition was found in model 2. Overall, the participants' ratings for the experimental items did not differ significantly from their ratings for the control items ($M_{change} = 1.629$, $p > .05$). Model 3 revealed a main effect of ObjectGender, such that all the participants' ratings for male objects were significantly lower than they were for female objects ($M_{change} = -8.621$, $p < .05$). Model 4 was a significant improvement over model 1 (χ^2 difference(4) = 189, $p < .001$) and had a smaller Akaike information criterion value, which indicates a better model (Hoffman & Rovine, 2007). In other words, adding an interaction effect between LanguageGroup and ObjectGender is more suitable for the pure main effects of LanguageGroup. Model 5 was not a significant

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Table 1. Descriptive statistics

							Skewness		Kurtosis	
Group	*N*	Min	Max	*M*	SE	SD	Statistic	Std. error	Statistic	Std. error
English monolingual										
Ratings for masculine objects	35	507.0	855.0	716.3	12.9	76.5	−.4	.4	.3	.8
Ratings for feminine objects	35	683.0	954.0	838.7	11.2	66.5	−.1	.4	.1	.8
Ratings for experimental objects	35	860.0	1,264.0	1,086.8	17.3	102.0	−.4	.4	.1	.8
Ratings for control objects	35	351.0	548.0	458.1	9.2	54.6	−.1	.4	−.4	.8
Ratings for all objects	35	26,976.0	671,667.0	385,639.3	29,292.5	173,296.8	.1	.4	−.4	.8
French monolinguals										
Ratings for masculine objects	35	634.0	1,268.0	933.9	27.6	163.0	.5	.4	−.5	.8
Ratings for feminine objects	35	393.0	909.0	704.6	20.7	122.3	−.5	.4	.0	.8
Ratings for experimental objects	35	849.0	1,379.0	1,125.1	23.3	137.8	.0	.4	−.4	.8
Ratings for control objects	35	428.0	614.0	523.3	10.2	60.5	.1	.4	−1.1	.8
Ratings for all objects	35	15,988.0	716,105.0	363,141.3	30,097.2	178,057.7	−.1	.4	−.3	.8
L1 English L2 French										
Ratings for masculine objects	35	555.0	1,316.0	884.8	38.1	225.7	.6	.4	−.7	.8
Ratings for feminine objects	35	367.0	938.0	716.1	27.4	161.9	−.9	.4	.0	.8
Ratings for experimental objects	35	949.0	1,252.0	1,112.9	15.6	92.3	.1	.4	−.9	.8
Ratings for control objects	35	324.0	607.0	485.4	12.2	72.3	−.3	.4	.0	.8
Ratings for all objects	35	12,302.0	803,857.0	329,500.2	39,598.6	234,268.3	.8	.4	−.3	.8
L1 French L2 English										
Ratings for masculine objects	35	643.0	1,250.0	897.2	27.0	159.6	.8	.4	.1	.8
Ratings for feminine objects	35	452.0	846.0	711.9	18.0	106.3	−.8	.4	.3	.8
Ratings for experimental objects	35	872.0	1,309.0	1,104.6	18.3	108.4	.1	.4	−.4	.8
Ratings for control objects	35	393.0	643.0	508.4	11.4	67.3	.3	.4	−.3	.8
Ratings for all objects	35	2,2261.0	989,887.0	314,549.1	44,305.9	262,117.4	.9	.4	.0	.8

Table 2. Results for main and interaction effects

	Model 1			Model 2			Model 3		
Parameter	Est.	SE	*t* value	Est.	SE	*t* value	Est.	SE	*t* value
Fixed effects									
(Intercept-English monolinguals)	47.6	2.1	22.7*						
(Intercept-conditionControl)				48.4	3.4	14.2*			
(Intercept-genderFeminine)							53.8	2.5	21.9*
French monolinguals	4.4	1.5	2.9*						
English-dominant bilinguals	1.3	1.5	.9						
French-dominant bilinguals	1.8	1.5	1.2						
conditionExperimental				1.6	4.1	.4			
genderMasculine							−8.6	3.4	−2.5*
Random effects									
Random object variance (SD)	106.6 (10.3)			109.7 (10.4)			90.5 (9.5)		
Random participants variance (SD)	21.0 (4.5)			22.7 (4.7)			22.7 (4.7)		
Residual variance (SD)	540.3 (23.2)			540.3 (23.2)			540.3 (23.2)		
Fit statistics									
Maximum-likelihood deviance (number of parameters)	41,118 (7)			41,127 (5)			41,121 (5)		
Akaike information criterion	41,132			41,137			41,131		
	Model 4			Model 5			Model 6		
Parameter	Est.	SE	t value	Est.	SE	t value	Est.	SE	*t* value
Fixed effects									
(Intercept-English monolinguals; genderFeminine)	43.8	2.7	16.3*						
(Intercept-English monolinguals; conditionControl)				45.4	3.6	12.5*			
(Intercept-English monolinguals; genderFeminine; conditionControl							48.8	4.6	10.5*
French monolinguals	14.9	1.8	8.5*	6.3	2.1	3.1*	14.7	2.6	5.5*
English-dominant bilinguals	11.8	1.8	6.7*	2.1	2.1	1.1	6.8	2.6	2.5*
French-dominant bilinguals	12.8	1.8	7.3*	3.2	2.1	1.6	8.7	2.6	3.2*
genderMasculine	7.4	3.6	2.1*				−6.8	6.4	−1.1
French monolinguals: genderMasculine	−21.1	1.9	−11.1*				−16.9	3.4	−4.9*
English-dominant bilinguals: genderMasculine	−21.1	1.9	−10.9*				−9.4	3.4	−2.7*
French-dominant bilinguals: genderMasculine	−22.1	1.9	−11.5*				−10.9	3.4	−3.1*
conditionExperimental				3.1	4.3	.7	−7.2	5.5	−1.3
French monolinguals: conditionExperimental				−2.8	2.1	−1.3	.2	2.9	.1
English-dominant bilinguals: conditionExperimental				−1.2	2.1	−.6	7.3	2.9	2.5*
French-dominant bilinguals: conditionExperimental				−2.1	2.1	−1.0	5.9	2.9	2.1*
genderMasculine:conditionExperimental							20.7	2.6	2.6*
French monolinguals: genderMasculine: conditionExperimental							−6.2	4.1	−1.4
English-dominant bilinguals: genderMasculine: conditionExperimental							−17.1	4.1	−4.1*
French-dominant bilinguals: genderMasculine: conditionExperimental							−16.1	4.1	−3.8*
Random effects									

(*Continued*)

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Table 2. (*Continued.*)

	Model 4			Model 5			Model 6		
Parameter	Est.	SE	*t* value	Est.	SE	*t* value	Est.	SE	*t* value
Random object variance (SD)	90.6 (9.5)			109.7 (10.4)			89.5 (9.5)		
Random participants variance (SD)	21.7 (4.6)			21.1 (4.6)			21.8 (4.7)		
Residual variance (SD)	518.1 (22.7)			540.4 (23.2)			515.8 (22.7)		
Fit statistics									
Maximum-likelihood deviance (number of parameters)	40,929 (11)*			41,116 (11)			40,901 (19)*		
Akaike information criterion	40,951			41,138			40,939		

improvement over model 1 (χ^2 difference(4) = 2.19, $p > .05$). This finding indicates that the interaction effect between LanguageGroup and Condition may not be as strong as the interaction between LanguageGroup and ObjectGender. Model 6 was another significant improvement over models 4 and 5 (χ^2 difference(8) = 28, $p < .001$; χ^2 difference(8) = 215, $p < .001$). This observation suggests that model 6, with the three-way interaction effect, was the best model to explain the data.

3.2 Demographic variables

A multiple regression analysis was conducted to see if the demographic variables (education level, participants' own gender, age, metalinguistic awareness (the ability to consciously reflect on the form and use of languages, Bialystok, 2001), L2 proficiency, self-reported L2 proficiency, L2 starting age, L2 exposure and language qualification) would be significant predictors of the voice ratings of all groups. The L2 proficiency and L2 exposure were focal variables in the regression because of the absence of proficiency tests in earlier literature. As a result, the regression consisted of two models. In model 1, participants' education level, gender, age, metalinguistic awareness, self-reported L2 proficiency, L2 starting age and language qualification were entered. Model 2 further included L2 proficiency and L2 exposure to see if model 2 could explain more variances in the outcome variables.

Regression analyses showed that model 1 did not explain a significant proportion of variances ($\Delta R^2 = .6\%$, $p > .05$). Model 2 only explained slightly more variances than model 1 ($\Delta R^2 = .1\%$, $p > .05$). Neither of the models could explain the variances significantly ($p > .05$). In other words, none of the demographic variables were significant predictors of participants' responses.

The proficiency tests in this study were crucial because significant differences were found between participants' self-perceived L2 proficiency and their actual L2 proficiency. Two Wilcoxon signed-rank tests indicate that English-dominant bilinguals' self-perceived L2 French proficiency was significantly higher than their tested French proficiency ($Z = -4.756$, $p < .001$). By contrast, French-dominant bilinguals' self-perceived English proficiency was significantly lower than their actual capabilities ($Z = -2.0$, $p < .05$).

Another intriguing issue is whether simultaneous and sequential bilinguals differ in their ratings. As most sequential bilinguals recruited in this study did not start learning L2 until ten, they could potentially show different ratings compared to simultaneous ones. Another several independent samples *t*-tests were conducted for the English-dominant and French-dominant bilingual groups. Results showed that sequential and simultaneous bilinguals did not differ significantly in their ratings for each Condition × ObjectGender combination.

3.3 English monolinguals versus French monolinguals

The ratings of 35 English and 35 French monolinguals for objects of four Condition × ObjectGender trials are presented in Figure 2.

The comparison between English and French monolinguals was conducted based on the most appropriate model, model 6, with the three-way interaction. Four independent sample *t*-tests were conducted to compare the ratings in each Condition (2) × ObjectGender (2) trial, as shown in Table 3.

The results of the independent sample *t*-tests indicated that there was a significant difference in the English and French monolingual participants' ratings for the experimental female

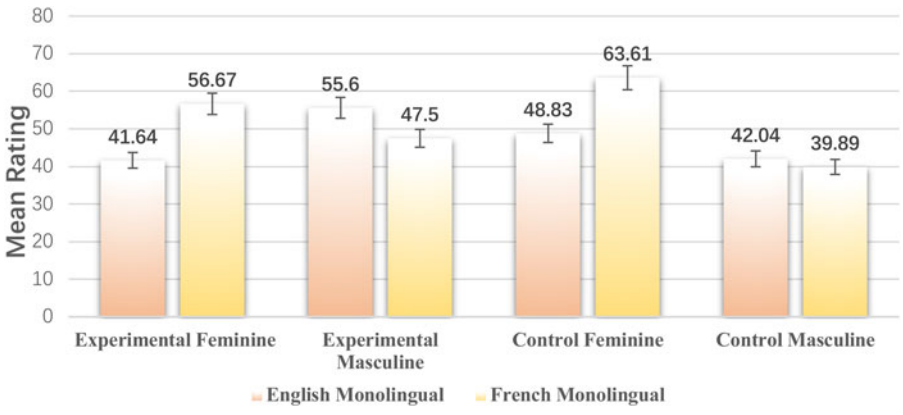

Figure 2. English monolinguals versus French monolinguals.

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Table 3. *t*-Test results for English and French monolinguals

Trial	Group	*M*	SD	SEM	df	*t*	*p* (two-tailed)	95% CI	*d*
Experimental × female	Eng. mono.	41.64	20.80	1.06	731	−8.8	<.01	−18.4, −11.7	−.6
	Fre. mono.	56.67	26.16	1.33					
Experimental × male	Eng. mono.	55.60	21.65	1.10	753	4.8	<.01	4.8, 11.4	.3
	Fre. mono.	47.50	24.94	1.27					
Control × female	Eng. mono.	48.83	22.14	1.67	340	−5.8	<.01	−19.8, −9.7	−.6
	Fre. mono.	63.61	25.73	1.95					
Control × male	Eng. mono.	42.04	24.19	1.89	348	.8	.41	−3.0, 7.3	.1
	Fre. mono.	39.89	24.96	1.89					

items: $t(731) = -8.8$, $p < .01$. The French monolinguals rated the experimental female items significantly[4] higher ($M = 56.67$, SD = 26.16) than did the English monolinguals ($M = 41.64$, SD = 20.80). Similarly, there was a significant difference for experimental male items between the two groups: $t(753) = 4.8$, $p < .01$. The French monolinguals also rated experimental male items significantly lower ($M = 47.50$, SD = 24.94) than did the English monolingual participants ($M = 55.60$, SD = 21.65).

With regards to the control items, there was a significant difference for control female items: $t(340) = -5.8$, $p < .01$. French monolinguals rated control female items significantly higher ($M = 63.61$, SD = 25.73) than did English monolinguals ($M = 48.83$, SD = 22.14). However, no significant difference was found for control male items: $t(348) = .80$, $p > .05$.

A follow-up chi-squared test was conducted to determine whether the groups' open-ended answers regarding the ten items rated highly differently by English and French monolinguals in the pilot study differed. A significant association was found between the groups and the answers: $\chi^2(2) = 92.70$, $p < .01$. French monolinguals gave answers that mapped to the language's grammatical gender system significantly more often ($N = 82$) than did English monolinguals ($N = 2$). English monolinguals relied more on the objects' characteristics to rate their genders ($N = 307$) than French monolinguals ($N = 217$).

3.4 English monolinguals versus English-dominant bilinguals

The ratings of 35 English and 35 English-dominant bilinguals for objects of four Condition × ObjectGender trials are presented in Figure 3.

Based on model 6, four independent sample *t*-tests were conducted (with a Bonferroni correction applied) to compare the ratings in each Condition (2) × ObjectGender (2) trial, as shown in Table 4.

The results of the independent sample *t*-tests indicated that there was a significant difference in the ratings of the English monolinguals and the English-dominant bilinguals for experimental female items: $t(720) = -8.10$, $p < .01$. In other words, the English-dominant bilinguals rated experimental female items significantly higher, i.e., more feminine ($M = 55.81$, SD = 27.07) than did the English monolinguals ($M = 41.64$, SD = 20.80). Similarly, there was a significant difference for the experimental male items between the two groups: $t(748) = 7.20$, $p < .01$. The English-dominant bilinguals also rated experimental male items significantly lower ($M = 43.29$, SD = 25.56) than did the English monolinguals ($M = 55.60$, SD = 21.65).

Regarding the control items, there was a significant difference for the control female items: $t(340) = -2.70$, $p < .01$. The English-dominant bilinguals rated control female items significantly higher ($M = 55.71$, SD = 25.77) than did the English monolinguals ($M = 48.83$, SD = 22.14). No significant difference was found for control male items: $t(348) = 1.0$, $p > .05$. The findings were consistent with those for the French monolinguals.

Follow-up correlation tests revealed no significant association between English-dominant bilinguals' French proficiency and their ratings for experimental female ($r = .36$, $p > .05$), experimental male ($r = .16$, $p > .05$) or control female items ($r = .34$, $p > .05$). Similarly, no significant association was found between ratings and L2 exposure, nor between ratings and metalinguistic awareness.

A follow-up chi-squared test revealed a significant association between the groups and their answers: $\chi^2(2) = 88.9$, $p < .001$. Like

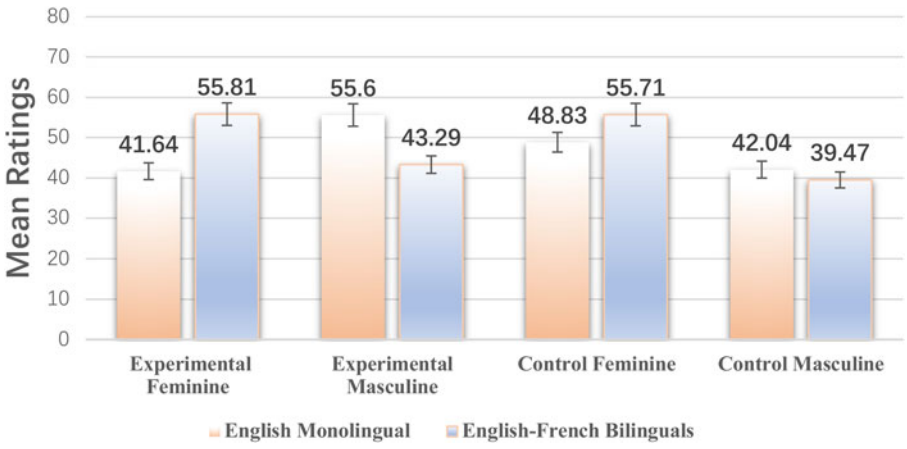

Figure 3. English monolinguals versus English-dominant bilinguals.

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Table 4. *t*-Test results for English monolinguals and English-dominant bilinguals

Trial	Group	*M*	SD	SEM	df	*t*	*p* (two-tailed)	95% CI	*d*
Experimental × female	Eng. mono.	41.64	20.80	1.06	720	−8.1	<.01	−17.6, −10.8	−.6
	Eng-dom. bil.	55.81	27.07	1.38					
Experimental × male	Eng. mono.	55.60	21.65	1.10	748	7.2	<.01	8.9, 15.7	.5
	Eng-dom. bil.	43.29	25.56	1.30					
Control × female	Eng. mono.	48.83	22.14	1.67	340	−2.7	<.01	−11.9, −1.8	−.3
	Eng-dom. bil.	55.71	25.77	1.95					
Control × male	Eng. mono.	42.04	24.19	1.83	348	1.0	.33	−2.6, 7.8	.1
	Eng-dom. bil.	39.47	25.32	1.91					

the French monolinguals, the English-dominant bilinguals answered significantly more in accord with the language's grammatical gender system ($N = 83$) than the English monolinguals ($N = 2$). The English monolinguals relied more ($N = 307$) on the objects' characteristics to rate their genders than the French monolinguals ($N = 228$).

3.5 French monolinguals versus French-dominant bilinguals

The ratings of 35 French monolinguals and 35 French-dominant bilinguals for objects of four Condition × ObjectGender trials are presented in Figure 4.

Based on model 6, four independent sample *t*-tests were conducted (with a Bonferroni correction applied) to compare the ratings in each Condition (2) × ObjectGender (2) trial, as shown in Table 5.

The results of the independent sample *t*-tests indicated that there was no significant difference in the ratings for experimental female objects: $t(768) = .20$, $p > .05$. In other words, French-dominant bilinguals rated experimental female items similarly to French monolinguals ($M = 56.31$, SD = 26.10 and $M = 56.67$, SD = 26.16, respectively). There was also no significant difference for experimental male items between the two groups after the Bonferroni correction was employed: $t(768) = 2.4$, $p = .02 > .0125$. French-dominant bilinguals only rated experimental male items insignificantly lower ($M = 43.18$, SD = 24.62) than did French monolinguals ($M = 47.50$, SD = 24.94). Regarding the control items, there was no significant difference for control female items: $t(348) = 2.1$, $p = .03 > .0125$. French-dominant bilinguals rated control female items insignificantly lower ($M = 57.58$, SD = 27.09) than did French monolinguals ($M = 63.61$, SD = 25.73). No significant difference was found for control male items: $t(348) = 1.0$, $p > .05$.

No significant association was found between French-dominant bilinguals' French proficiency and their ratings ($r = -.18$, $p > .05$). Other demographic variables (metalinguistic awareness, L2 exposure, starting age of L2) were not significantly correlated with the ratings.

A follow-up chi-squared test revealed no significant relationship between the groups and the answers: $\chi^2(2) = 1.4$, $p > .05$. French monolinguals and French-dominant bilinguals produced similar answer patterns regarding the grammatical gender system ($N = 82$; $N = 76$), objects' characteristics ($N = 217$; $N = 231$) and other features ($N = 51$; $N = 43$).

4. Discussion

This study investigated linguistic relativity hypothesis, specifically, the effect of the grammatical gender system of French (both as L1 and L2) on its speakers' and learners' perceptions of objects. Firstly, the study investigated if (and if so, how) the grammatical gender system in French affects French monolinguals compared with English monolingual speakers. Secondly, it explored if (and if so, how) learning a gendered or genderless L2 affects bilingual adults' perception of objects compared to monolingual speakers. Thirdly, it tested if the patterns of object perception between

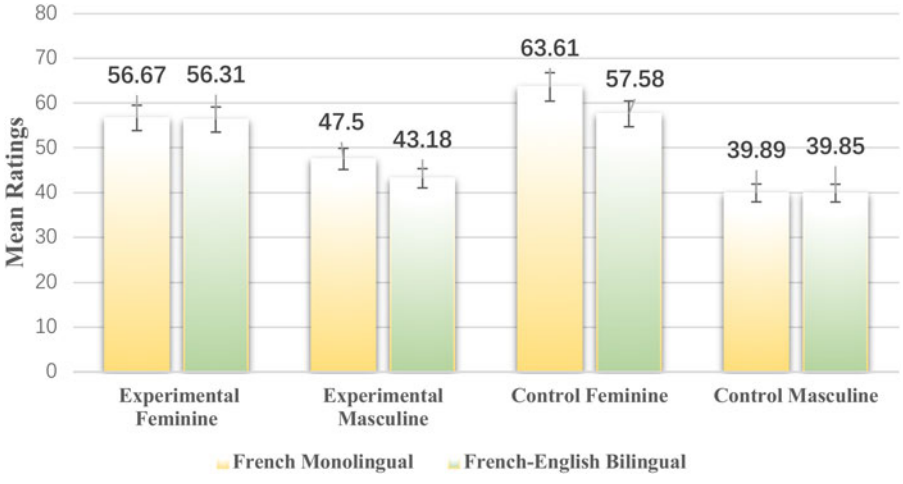

Figure 4. French monolinguals versus French-dominant bilinguals.

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Table 5. *t*-Test results for French monolinguals and French-dominant bilinguals

Trial	Group	*M*	SD	SEM	df	*t*	*p* (two-tailed)	95% CI	*d*
Experimental × female	Fre. mono.	56.67	26.16	1.33	768	.2	.85	−3.3, 4.1	.01
	Fre.-dom. bil.	56.31	26.10	1.33					
Experimental × male	Fre. mono.	47.50	24.94	1.27	768	2.4	.02	.8, 7.8	.2
	Fre.-dom. bil.	43.18	24.62	1.25					
Control × female	Fre. mono.	63.61	25.73	1.95	348	2.1	.03	.5, 11.6	.2
	Fre.-dom. bil.	57.58	27.10	2.05					
Control × male	Fre. mono.	39.89	24.96	1.89	348	1.0	.99	−5.2, 5.3	.0
	Fre.-dom. bil.	39.85	25.02	1.89					

English-dominant bilingual adults and French-dominant bilingual adults are different and, if so, how.

The study's findings showed an effect of the French grammatical gender system on object perception. The results indicated that the French monolinguals and the English-dominant bilinguals assigned objects' gender consistently with the French grammatical gender system. Secondly, the effect of French grammatical gender was not significantly affected by the acquisition of a genderless L2 (English). This was showcased in the result that French-dominant bilinguals did not differ significantly from French monolinguals in ratings, though having lower ratings for experimental masculine and control feminine objects in descriptive statistics.

4.1 Effect of the French grammatical gender system on perception

The results of this experiment support linguistic relativity in that an effect of the French grammatical gender system did exist. Also, because there was no significant difference in sequential and simultaneous bilinguals in ratings, the effect can be argued to hold across the bilingualism spectrum (cf. Kurinski & Sera, 2011). Importantly, the effect of French as an L2 on perception can only be found among speakers with a genderless L1.

Such an effect of the French grammatical gender system can have practical societal implications. For example, children's literature has been found to present anthropomorphised objects or animals with genders following the grammatical gender system (Bassetti & Nicoladis, 2016). These gendered figures could influence children's perceptions and impose gender biases from a very young age. This idea was supported by the responses of the French monolinguals and French-dominant bilinguals in the experiment's explanation questions. When assigning gender to *elephant*, most French participants considered it masculine, following the elephant's gender in French. However, four participants explained that their score was motivated by recalling Babar, a fictional elephant character in the French children's book *Babar the Elephant*. This explanation shows the long-lasting effect of character depictions in children's literature on their perception.

Such an effect also appears in children's name production. A recent study conducted by Hsiao et al. (2021) investigated the personal name usage in children's literature, and stories written by over 100,000 children aged 5–13 in the UK. Results showed that male authors overrepresent male names in children's books: male authors produced 159,579 names, of which 71% (112,998) were male names. The other finding was that boys wrote more about boys with age due to the input they were exposed to: among 429,804 names produced by boys, 85% ($N = 363,189$) were male names, and only the remaining 15% ($N = 66,615$) were female. By contrast, female authors and little girls used more balanced names in their stories. The findings suggest that the written language in children's books can shape and be shaped by gender biases (Hsiao et al., 2021).

4.2 Genderless L2 does not weaken the effect of French grammatical gender

In accord with predictions to RQ3, the current study found no significant effect, except a marginal difference in descriptive statistics. Thus, acquiring a genderless L2 might not effectively reduce the effect of the original L1 grammatical gender system. This notion resonates strongly with earlier studies (Bassetti & Nicoladis, 2016; Boroditsky & Schmidt, 2000; Forbes et al., 2008).

Additionally, in this study, most French-dominant bilinguals had an advanced level of English proficiency. It is thus possible to raise that the effect of the French gender system would not be affected by even a high level of English proficiency. This outcome is contrary to that of Sato et al. (2013), who found the effect of the French gender system to be diminished by a high level of English proficiency in French–English bilinguals. The discrepancy could be due to the different focus of comparison of the study. Sato et al. (2013) compared English–French bilinguals' responses with French–English bilinguals, not French monolinguals, as in this study. Resemblances between English–French bilinguals' responses and French–English bilinguals can hardly lead to the conclusion that acquiring English had an effect. The lack of French monolinguals as a baseline for comparison could be an issue with Sato et al.'s (2013) study.

4.3 Independence from L2 proficiency and other demographic variables

Another important finding was that a higher level of proficiency in a gendered L2 did not indicate a more potent effect of the gender system on perception; further, neither did a higher level of proficiency in a genderless L2 lead to a weaker impact of the L1 grammatical gender system. These findings coincide with Kurinski and Sera's (2011) and strengthen its analysis by adding the L2 proficiency test. A note of caution is due here since literature shows that, for speakers of a gendered L1, a higher proficiency in another gendered L2 would weaken the effect of the L1 grammatical gender system (e.g., Bassetti & Nicoladis, 2016).

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Therefore, the combination of bilingual or multilingual languages is important when interpreting the role of L2 proficiency in the effect of the L1 gender system.

Similarly, for speakers of a gendered L1, the impact of a gendered L1 may potentially remain relatively high with an increase in genderless L2 proficiency. However, this finding is contrary to Sato et al. (2013), who found that an increase in English proficiency decreases the effect of the L1 French gender system. The different usage of tasks to measure perception could cause such a difference. This study used a voice distribution task, whereas Sato et al. (2013) adopted the rarely used sensibility judgement task (Samuel et al., 2019). The voice distribution task has been found likely to yield supporting results of linguistic relativity because its instruction explicitly contains gender (Samuel et al., 2019). Thus, it might increase the chances of participants unconsciously referring to the grammatical gender system to complete the task (Bender et al., 2011). For example, when justifying why s/he gave a feminine rating to *table*, an English monolingual participant in this study mentioned the French grammatical gender system. Perhaps, the consent sheet in both languages and the instruction of the voice task influenced the participant. However, Sato et al.'s (2013) sensibility judgement task also has limitations in emphasising stereotypes (e.g., the association between *social workers* and feminity). These limitations for both tasks resonate with a criticism of this line of linguistic relativity research: results depend heavily on the task types (Samuel et al., 2019). There is ample room for further progress in determining the role of genderless L2 proficiency. Adopting a neurolinguistic approach, using different tasks, and proper forms of L2 proficiency testing are helpful.

The proficiency measure turned out to be crucial. A significant discrepancy was found between English participants' self-perceived higher French proficiency and their lower capability. On the contrary, French-dominant bilinguals tended to underestimate their English proficiency. Such a Dunning–Kruger effect contrast resonates with an earlier study conducted by Trofimovich et al. (2016), in which high-proficient learners underestimated their L2 abilities while less-proficient learners overestimated them. These findings consider that earlier studies using participants' self-perceived proficiency as the measure for L2 proficiency (e.g., Lambelet, 2016) may lack validity to a certain extent, rendering their results less convincing.

Apart from L2 proficiency, other demographic variables (metalinguistic awareness, L2 exposure, starting age of L2) were found not to predict the ratings for objects. The self-reported data for metalinguistic awareness and L2 exposure may limit the interpretation (Dörnyei & Taguchi, 2009). It may also be possible that metalinguistic awareness affects how participants learn the grammatical gender system (e.g., French words ending with consonants are usually masculine) (Brooks & Kempe, 2012), but not how the gender system creeps into perception. Regarding simultaneous and sequential bilinguals, the study did not find difference between these two kinds of bilinguals, which might be because of the small sample size of simultaneous bilinguals (only one simultaneous French-dominant bilingual was found). The unequal sample size can limit the rigour of data analysis (Shaw & Mitchell-Olds, 1993).

4.4 French might introduce gender stereotypes

An unexpected finding of the study was that the results of the comparison of four levels all (somewhat) clustered in control, feminine items. Although the comparison of French monolinguals and French-dominant bilinguals did not give rise to a significant difference, the descriptive statistics showed a slight difference in the two groups' ratings. Control items in this study were carefully chosen based on the piloting; they had to be the objects associated with typical gender biases. A significantly higher gender rating in control female objects could potentially suggest that acquiring French as an L2 might start thinking of 'tables' as feminine due to their exposure to French.

The acquisition of English as an L2 might not help reduce the gender biases from the initial effect of L1 French. RQ3 showed that French-dominant bilinguals rated control feminine items slightly lower than French monolinguals. Future research can be undertaken with more language pairs to see if similar effects emerge. Future research can also focus on what kind of gender biases would be introduced by French, as only feminine gender biases were shown in this study (none of the four groups differed significantly in ratings for the control male items).

4.5 Pedagogical implications and future of linguistic relativity research

Teachers might need to be aware of the effect of the grammatical gender system when teaching an L2 with a gender system. Students could perceive objects as masculine or feminine depending on their L1/L2 and use the wrong grammatical gender in production. Teachers can thus emphasise the discrepancies between perception and grammar during teaching. Emphasising the difference could be important because studies have shown that speakers with a genderless L1 often find the L2 grammatical gender system challenging to acquire (e.g., Sabourin et al., 2006; Shimanskaya & Slabakova, 2019). Also, when speakers perceive an object based on their L1 that is different from its true grammatical gender in the L2, cross-linguistic influence (CLI) might come into play. Nicoladis et al. (2021) touched on this by investigating whether the subconscious intuitions that native English speakers have about objects' gender could impact English learners of French in judging the French object's gender. Results showed that participants were more accurate with French words with congruent English gender connotations. This finding indicates the influence of English on the acquisition of the French gender system (Nicoladis et al., 2021).

CLI might complement the role of linguistic relativity in affecting the gender system. Participants with different L1s might think of the objects' French gender differently depending on L1 and/or L2 influences. The effect of CLI might skew the (interpretation of) results of linguistic relativity research. Thus, the two phenomena can be combined, bringing more pedagogical implications for CLI and linguistic relativity research.

4.6 Limitations

Despite the interesting results it gave rise to, the present study was not without limitations. Firstly, although LexTALE was an objective L2 proficiency test, it can only offer an indication of the vocabulary aspect of L2 proficiency (de Bruin, 2019). Secondly, the voice distribution task cannot fully reflect people's perceptions as it tended to yield more supportive evidence for the effect of the grammatical gender system on perception (Samuel et al., 2019). In contrast, the properties judgement task (in which participants were asked to describe the properties or characteristics) tended to have more disapproving evidence against the same effect.

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press

Although the latter properties judgement task might be partially biased because it usually contains gender stereotypes (properties like 'tough' associated with masculinity and 'soft' with feminity), future studies can utilise both tasks to add more rigour. Whether participants would give the same answer patterns in these different methods can also be an intriguing direction.

Moreover, earlier studies noted that participants' perceptions might be influenced by whether the objects were animate things or artefacts (e.g., Bassetti, 2007). Thus, future research can choose objects more selectively by balancing the number of (non-)artefacts or (in-)animate items.

An unavoidable limitation of this study is that only adults were recruited. However, as discussed in the literature review, there was a potential age boundary (seven or eight), after which the language's effect started to play a role (Flaherty, 2001). Therefore, future research would be beneficial to investigate when these effects of the gender system on perception begin to emerge. Such a direction has important theoretical implications regarding the debate on linguistic determinism and universalism. Neither of the two strong accounts could hold if language influences thought after a certain age.

5. Conclusion

The linguistic relativity hypothesis is supported by the current findings: the French grammatical gender system seems to creep into the minds of its monolingual speakers and learners. With the objective L2 proficiency measurements, the findings also contribute to the validation of previous arguments that a genderless L2 (even at a high level of proficiency) cannot impact the effect of the gendered L1 on perception.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1366728924000464.

Data availability statement. The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements. We thank everyone who participated in the data collection and those who helped the authors in the research.

Competing interests. The authors declare that they have no competing interests.

Notes

1 English–French means dominant in English (i.e., English was the L1); French–English means the opposite; 14% of the participants were simultaneous bilinguals.

2 Note that an effect size of .29 was used based on Bassetti (2007), discussed in detail in the previous section.

3 The B2 level refers to the Common European Framework of Reference for Languages proficiency level. At the B2 level, learners are considered to have an upper-intermediate level of proficiency, whereas B1 indicates 'intermediate', C1 indicates 'advanced' and C2 indicates 'near-native'.

4 The independent samples' *t*-tests were still significant after the Bonferroni correction was applied ($p < .0125$).

References

Ahearn, L. M. (2017). *Living language: An introduction to linguistic anthropology* (2nd ed.). Wiley Blackwell.

Almutrafi, F. (2015). Language and cognition: Effects of grammatical gender on the categorisation of objects (Order No. 10304012) [Doctoral dissertation, University of Newcastle]. ProQuest Dissertations and Theses Global.

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism*, *12*(1), 83–95. doi: 10.1017/S136672890800388X

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412. doi: 10.1080/02643294.2020.1769050

Bassetti, B. (2007). Bilingualism and thought: Grammatical gender and concepts of objects in Italian–German bilingual children. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behaviour*, *11*(3), 251–273. doi: 10.1177/13670069070110030101

Bassetti, B., & Nicoladis, E. (2016). Research on grammatical gender and thought in early and emergent bilinguals. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behaviour*, *20*(1), 3–16. doi: 10.1177/1367006915576824

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bender, A., Beller, S., & Klauer, K. C. (2011). Grammatical gender in German: A case for linguistic relativity? *Quarterly Journal of Experimental Psychology*, *64*(9), 1821–1835. doi: 10.1080/17470218.2011.582128

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

Bickel, B. (2000). Grammar and social practice: On the role of 'culture' in linguistic relativity. In S. Niemeier & R. Dirven (Eds.), *Amsterdam Studies in The Theory and History of Linguistic Science Series, 4* (pp. 161–192). John Benjamins Publishing Company.

Boroditsky, L., & Schmidt, L. A. (2000). Sex, syntax, and semantics. In L. R. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the 22nd conference of the cognitive science society* (pp. 42–47). Lawrence Erlbaum Associates.

Brooks, P. J., & Kempe, V. (2012). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition*, *41*(2), 281–296. doi: 10.3758/s13421-012-0262-9

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, *49*(3), 454–462. doi: 10.1037/h0057814

Brysbaert, M. (2013). LexTALE_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, *53*(1), 23–37. doi: 10.5334/pb-53-1-23

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, *58*(1), 63–79. doi: 10.1111/j.1467-9922.2008.00462.x

Chen, Z. (2020). Language's effect on cognitive ability: A glimpse into the conspiracy theories. *Advances in Educational Technology and Psychology*, *16*, 215–255.

Clarke, M. A., Losoff, A., McCracken, M. D., & Still, J. (1981). Gender perception in Arabic and English. *Language Learning*, *31*(1), 159–169. doi: 10.1111/j.1467-1770.1981.tb01377.x

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioural Science*, *9*(3), 33–46. doi: 10.3390/bs9030033

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi: 10.3758/BF0319314

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *The Behavioral and Brain Sciences*, *39*, e229. doi: 10.1017/S0140525X15000965

Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31. doi: 10.1177/0022022101032001005

Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015). On the road to somewhere: Brain potentials reflect language effects on motion

event perception. *Cognition*, *141*, 41–51. doi: 10.1016/j.cognition.2015.04.006

Forbes, J. N., Poulin-Dubois, D., Rivero, M. R., & Sera, M. D. (2008). Grammatical gender affects bilinguals' conceptual gender: Implications for linguistic relativity and decision making. *The Open Applied Linguistics Journal*, *1*(1), 68–76.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. doi: 10.3758/BF03192848

Hsiao, Y., Banerji, N., & Nation, K. (2021). Boys write about boys: Androcentrism in children's reading experience and its emergence in children's own writing. *Child Development*, *92*(6), 2194–2204. doi: 10.1111/cdev.13623

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. doi: 10.1177/0956797611432178

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, *22*(5), 519–534. doi: 10.1007/BF01068252

Kuo, J. Y. C., & Sera, M. D. (2009). Classifier effects on human categorization: The role of shape classifiers in mandarin Chinese. *Journal of East Asian Linguistics*, *18*(1), 1–19. doi: 10.1007/s10831-008-9036-6

Kurinski, E., & Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism*, *14*(2), 203–220. doi: 10.1017/S1366728910000179

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lambelet, A. (2016). Second grammatical gender system and grammatical gender-linked connotations in adult emergent bilinguals with French as a second language. *The International Journal of Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *20*(1), 62–75. doi: 10.1177/1367006915576832

Lemhöfer, K. M., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. doi: 10.3758/s13428-011-0146-0

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, *56*(1), 1–24. doi: 10.1515/iral-2017-0016

Nicoladis, E., & Foursha-Stevenson, C. (2012). Language and culture effects on gender classification of objects. *Journal of Cross-Cultural Psychology*, *43*(7), 1095–1109. doi: 10.1177/0022022111420144

Nicoladis, E., Westbury, C., & Foursha-Stevenson, C. (2021). English Speakers' implicit gender concepts influence their processing of French grammatical gender: Evidence for semantically mediated cross-linguistic influence. *Frontiers in Psychology*, *12*, 740920. doi: 10.3389/fpsyg.2021.740920

Pandey, A. (2017, May 3). The language you speak affects how your brain experiences the passage of time. *International Business Times*. https://www.ibtimes.com/language-you-speak-affects-how-your-brain-experiences-passage-time-2533953

Pavlenko, A. (2003). Eyewitness memory in late bilinguals: Evidence for discursive relativity. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *7*(3), 257–281. doi: 10.1177/13670069030070030301

Perlovsky, L. (2009). Language and emotions: Emotional Sapir–Whorf hypothesis. *Neural Networks*, *22*(5), 518–526. doi: 10.1016/j.neunet.2009.06.034

Pinker, S. (1994). *The language instinct*. W. Morrow and Co.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1–29. doi: 10.1191/0267658306sr259oa

Samuel, S., Cole, G., & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin and Review*, *26*(6), 1767–1786. doi: 10.3758/s13423-019-01652-3

Sapir, E. (1956). Language. In D. G. Mandelbaum (Ed.), *Culture, language and personality: Selected essays* (pp. 1–44). University of California Press.

Sato, S., & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: Evidence for the structural–feedback hypothesis. *Cognition*, *176*, 220–231. doi: 10.1016/j.cognition.2018.03.014

Sato, S., Gygax, P. M., & Gabriel, U. T. E. (2013). Gender inferences: Grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism*, *16*(4), 792–807. doi: 10.1017/S1366728912000739

Sera, M. D., Elieff, C., Forbes, J., Clark Burch, M., Rodríguez, W., & Poulin Dubois, D. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology*, *131*(3), 377–397. doi: 10.1037/0096-3445.131.3.377

Shaw, R. G., & Mitchell-Olds, T. (1993). ANOVA for unbalanced data: An overview. *Ecology*, *74*(6), 1638–1645. doi: 10.2307/1939922

Shimanskaya, E., & Slabakova, R. (2019). L1–L2 differences in the L2 classroom: Anticipating Anglophone learners' difficulties with French pronoun interpretation. *Language Teaching Research*, *23*(2), 259–277. doi: 10.1177/1362168817739650

Thompson, G., & Dooley, K. (2019). Ensuring translation fidelity in multilingual research. In J. McKinley, & H. Rose (Eds.). *The Routledge handbook of research methods in applied linguistics* (pp. 63–75). Routledge.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism*, *19*(1), 122–140. doi: 10.1017/S1366728914000832

Vernich, L., Argus, R., & Kamandulyte-Merfeldiene, L. (2017). Extending research on the influence of grammatical gender on object classification: A cross-linguistic study comparing Estonian, Italian and Lithuanian native speakers. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, *13*, 223–240. doi: 10.5128/ERYa13.14

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology*, *134*(4), 501–520. doi: 10.1037/0096-3445.134.4.501

Wasserman, B. D., & Weseley, A. J. (2009). Qué? Quoi? Do languages with grammatical gender promote sexist attitudes? *Sex Roles*, *61*(9–10), 634–643. doi: 10.1007/s11199-009-9696-3

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Zubin, D. A., & Köpcke, K. M. (1984). Affect classification in the German gender system. *Lingua. International Review of General Linguistics/Revue Internationale de Linguistique Generale*, *63*(1), 41–96. doi: 10.1016/0024-3841(84)90031-7

https://doi.org/10.1017/S1366728924000464 Published online by Cambridge University Press