Linguistic Relativity: Object Categorisation Differences Between Arabic and English Speakers

Abdullah Bin Dawood[1*]
Maya G. Sen[2]
Yan Jing Wu[3]

[1]*Department of Psychology, College of Education, King Saud University, Riyadh, Saud Arabia*

[2]*City College of San Francisco, California, USA*

[3]*College of Psychology and Sociology, Shenzhen University, Shenzhen, China*

*Corresponding e-mail:[aaldawood1@ksu.edu.sa]

We examined the potential effects of the Arabic grammatical gender system on object categorisation using an online voice attribution task. Compared to native English speakers (including English monolinguals and English-Arabic bilinguals), native Arabic speakers (including Arabic monolinguals and Arabic-English bilinguals) were more likely to assign either a man's voice (or a boy's voice) or a woman's voice (or a girl's voice) to inanimate objects with a gender that was consistent with the objects' grammatical gender in Arabic. Interestingly, when assigning genders to objects that do not have an associative stereotypical gender, a male-attribution tendency was found in both native Arabic speakers and native English speakers. Additionally, while native Arabic speakers and native English speakers assigned voices consistently with Arabic grammatical gender (GG) to objects with an associative stereotypical gender compatible with its grammatical gender in Arabic, they assigned voices to objects with an associative stereotypical gender incompatible with its grammatical gender in Arabic, likely based on object-gender stereotypical associations. Additionally, the performance of Arabic and English monolinguals was highly comparable with that of Arabic-English and English-Arabic bilinguals. We conclude that while the effects of linguistic structure on object categorisation might be generalised to Oriental languages, these effects are constrained and sometimes over-ridden by object-gender stereotypical associations, supporting a fully interactive account of the linguistic relativity hypothesis.

Keywords: bilingualism, linguistic relativity, voice attribution task, grammatical gender, categorisation

The linguistic relativity hypothesis refers to the idea that the linguistic structure of a language affects its speaker's view of the world (Garvin, 1958). While a strong version of its interpretation, which contends language determinism, has been challenged by evidence of language-independent cognitive operations (e.g., Chomsky, 2000; J. A. Fodor, 1975; Pinker, 1995), researchers such as Slobin (1996) have argued that language shapes the way individuals perceive reality and mediates their thinking at least for the moment of speaking (i.e., the "thinking for speaking" view). A number of studies have shown that lexical characteristics of languages restrict domain-general cognitive processes such as numerical cognition (Athanasopoulos, 2006; Brysbaert, Fias, & Noel, 1998; Gordon, 2004; Miller, Smith, Zhu, & Zhang, 1995; Moeller, Shaki, Göbel, & Nuerk, 2015), colour perception (Davidoff, Davies, & Roberson, 1999; Davies, 1998; He et al., 2019; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; Xia, Xu, & Mo, 2019; Özgen, 2004; Özgen & Davies, 2002), and spatial representation (Flecken, 2011; Majid, Bowerman, Kita, Haun, & Levinson, 2004; Papafragou & Selimis, 2010). For instance, in a seminal study, Gordon (2004) examined mathematical abilities of Piraha speakers, who use the 'one-two-many' counting system (the word 'many' refers to numbers 3 or larger). Results show that this innumerate counting system limits the performance of Piraha speakers when operating numerical tasks involving quantities greater than 3.

In the same vein, Sera, Berge del Castillo Pintado (1994) and Thierry et al. (2009) showed that the colour terminologies of a speaker's language affect his or her perception of colours at the pre-attentive level. Greek speakers dissociated light blue (*ghalazio*) and dark blue (*ble*) to a greater extent than English speakers, whereas no such a difference was observed in the light-dark green comparison (both Greek and English have only one word for 'green' regardless of its luminance). Language-specific terminology may affect behavioural performance through shaping the mental representation of entities such as objects, colours, and numbers. However, as terminologies are categorised in terms of their semantic content (e.g., boy and girl refer to a male and a female child, respectively, in English) (Sera et al., 1994), they may affect object categorisation through sub-vocal activities (e.g., implicit access to objects' names) during experimental tasks.

Unlike terminologies, grammar, which is defined as the group of rules of language use, does not necessarily carry specific semantic content. For example, in some languages (e.g., English and Chinese), third-person singular pronouns are marked for gender (for Chinese, only when written), which forms part of the meaning of the word. In other languages, such as Arabic and Hebrew, the grammatical gender (GG) system does not always indicate gender information (e.g., in the case of an inanimate concept). It may be used as a noun classification system that divides nouns into gendered categories and marks relevant parts in a sentence for gender agreement (Beit-Hallahmi et al., 1974).

Previous studies have shown GG effects on the perception and categorisation of objects (Andonova, Gosheva, Janyan, & Schaffai, 2007; B. A. Bassetti, 2014; Clarke, Losoff, McCracken, & Still,

1981; Ervin, 1962; Konishi, 1993; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). For instance, Clarke et al. (1981) asked Arabic and English speakers to rate concrete objects and abstract concepts that are essentially asexual on a masculine-feminine scale and found that the two groups categorised the same set of stimuli in significantly different ways. Arabic speakers rated asexual objects and concepts more consistently with their grammatical gender in Arabic as compared to English speakers. Moreover, the acquisition of a second language that has different GG rules from the native language has been shown to affect both the linguistic and conceptual representations of bilinguals (Andonova et al., 2007; B. Bassetti, 2007; B. A. Bassetti, 2014; Cubelli, Paolieri, Lotto, & Job, 2011; Kousta, Vinson, & Vigliocco, 2008; Nicoladis & Foursha-Stevenson, 2011). Andonova et al. (2007) found that native Bulgarian speakers with German as a second language were influenced by their second language GG when assigning genders to grammatrically neutral words in Bulgarian that are grammatically feminine or masculine in German.

In addition to categorisation tasks, a nonverbal procedure known as the voice attribution task (VAT) has been used to test the incidental effects of GG on participant's conceptual organisations (Forbes, Poulin-Dubois, Rivero, & Sera, 2008; Kurinski, Jambor, & Sera, 2016; Kurinski & Sera, 2011; Lambelet, 2016; Sera et al., 1994; Sera et al., 2002; Vernich, Argus, & Kamandulytė-Merfeldienė, 2017). In a classic VAT, the participant is presented with images of objects and asked to assign either a female or a male voice to them, usually in a context that orients the participant away from the GG of the pictures' names (e.g., making a cartoon film using these photos). Studies found that the gender of the voice attributed to inanimate objects by the participant is influenced by non-linguistic conceptual gender representation (e.g., natural objects are assigned more feminine voices while artificial objects are assigned more masculine voices) (Mullen, 1990; Vernich et al., 2017) and the GG system of the language that the participant speaks (Forbes et al., 2008; Kurinski, Jambor, & Sera, 2015; Sera et al., 1994; Sera et al., 2002). Strikingly, when testing bilingual participants who speak a gendered language (e.g., Spanish or French) and a genderless or nearly genderless language (e.g., Hungarian or English), their performance in the VAT is affected by the GG system of the gendered language regardless of whether it is their first or second language (Forbes et al., 2008). Furthermore, this effect increases as the proficiency level in the gendered language increases (Kurinski et al., 2015; Kurinski & Sera, 2011).

Previous studies examining GG effects on categorisation tasks have exclusively been conducted in laboratory contexts, where the sex of the experimenter (or task presenter) may exert a bias on participants' performance. For example, studies have found that participants are faster and more accurate in identifying the GG of an object's name pronounced by a gender-congruent (e.g., feminine word spoken by a female voice) speaker as compared to a gender-incongruent speaker. In addition, previous studies have also shown that the colours in which objects are presented affect participants' gender classification (Ellis & Ficek, 2001; Picariello, Greenberg, & Pillemer, 1990).

The current study investigates the effect of a grammatical gender system on the conceptual representation of objects by comparing the performance of native Arabic speakers with that of native English speakers using a VAT with uncoloured hand drawings of inanimate objects. Arabic and English vary based on what might be called a grammatical gender system. Arabic has a two-gender system in which almost every noun has either (marked) feminine or (default/unmarked) masculine gender (Alkohlani, 2016). The grammatical gender can be semantically driven in the case of the majority of biological entities but can also be arbitrarily driven in cases of inanimate entities. However, English has no grammatical gender, except for the third-person singular pronouns "he" and "she" (Beit-Hallahmi et al., 1974; Sera et al., 1994). The feminine-masculine gender system of Arabic makes the current study highly comparable to previous findings on languages with similar grammatical features (e.g., French, Italian, and Spanish) (Forbes et al., 2008; Kurinski et al., 2015; Sera et al., 1994; Sera et al., 2002).

To examine the independent processing of linguistic structures (e.g., GG) and non-linguistic conceptual gender representation (e.g., *Natural-feminine/Artificial-masculine*), the online VAT employs natural and artificial objects with different Arabic GG. Previous studies have shown that people are inclined to categorise natural objects as more feminine and artificial objects as more masculine (Mullen, 1990; Sera et al., 1994). Similar to the literature of linguistic relativity, the Natural-feminine/Artificial-masculine hypothesis has been mostly studied with speakers of Indo-European languages.

To investigate the potential interplay between GG and social factors (object-gender stereotypical association) during object categorisation (Nicoladis & Foursha-Stevenson, 2011), the VAT separates the consonant category, which consists of objects with socially/functionally congruent GG (e.g., "axe" is primarily associated with males and is masculine in Arabic), from the dissonant category, which consists of objects with socially/functionally incongruent GG (e.g., "dress" is primarily associated with females but is masculine in Arabic). A strong version of the linguistic relativity hypothesis (i.e., linguistic determinism) would predict that object categorisation in the dissonant category is either completely determined or affected by their Arabic GG for the native Arabic speakers regardless of the contradiction between the social factors and grammatical gender.

To investigate the impact of acquiring a second language on object categorisation, we also examine native Arabic speakers who have learned English as a second language and native English speakers who have learned Arabic as a second language. The linguistic relativity hypothesis would predict a different VAT performance of Arabic speakers from that of English speakers, whose gender assignment would be influenced by the GG of the pictures' names in Arabic. Based on previous findings, it is also predicted that both Arabic-English and English-Arabic bilingual speakers would demonstrate Arabic GG effect on their performance in the VAT (Forbes et al., 2008), with a stronger effect in the performance of Arabic-English bilinguals as compared to that of English-Arabic bilinguals (Kurinski et

al., 2015; Kurinski & Sera, 2011).

Method

Participants

One hundred and twenty participants, comprised of 30 Arabic monolinguals (13 female and 17 male), 30 English monolinguals (22 female and 8 male), 30 Arabic-English bilinguals (6 female and 24 male), and 30 English-Arabic bilinguals (8 female and 22 male) volunteered for the experiment, which was approved by the Northern Michigan University institutional review board for research involving human subjects. All participants had university-level education. Arabic monolinguals were Psychology undergraduates from King Saud University in Riyadh, KSA. English monolinguals were Psychology undergraduates from Northern Michigan University, USA. English-Arabic bilinguals were recruited from native English speakers who teach Arabic at American universities. Arabic-English bilinguals were recruited from the 'Lecturers and Assistant Teachers Association' (قـروب المعيـدين و المحاضـرين) on Facebook (https://www.facebook.com/groups/ksauniv/). Members of this association are native Arabic speakers who teach in American universities. All bilingual participants rated their level of proficiency in second language as either Beginner, Intermediate, or Advanced. In the Arabic-English bilingual group, 20 rated themselves as advanced bilinguals, 9 as intermediate, and 1 as beginner. In the English-Arabic bilingual group, 10 rated themselves as advanced, 14 as intermediate, and 6 as beginner.

Materials

Forty pictures of uncoloured hand-drawn objects were used in the current experiment to reduce any possible confounds due to variances in colours (Ellis & Ficek, 2001; Picariello et al., 1990). Eight pictures were used in the familiarisation phase to ensure that participants focused on sex attributes of the objects rather than the GG of their Arabic names. Half of these objects' Arabic names were feminine, and the other half were masculine, while their gender attributes were manipulated orthogonally (e.g., a female and a male doctor). The remaining thirty-two pictures were used as test stimuli and included four objects in each of the following categories' conditions: natural-feminine, natural-masculine, artificial-feminine, artificial-masculine, consonant-feminine, consonant-masculine, dissonant-feminine, and dissonant-masculine (see Table 1).

Table 1

Categories' Conditions and Words for Objects Used

NF	NM	AF	AM	CF	CM	DF	DM
Hand	Arm	Spoon	Bridge	Skirt	Saw	Knife	Dress
Star	Neck	Window	Phone	Bracelet	Drill	Airplane	Jewel
Ear	Body	Printer	Train	Flowers	Axe	Bomb	Perfume
Knee	Planet	Table	Book	Earrings	Screwdriver	Tank	Necklace

Labels used in the first row refer to natural-feminine (NF), natural-masculine (NM), artificial-feminine (AF), artificial-masculine (AM), consonant-feminine (CF), consonant-masculine (CM), dissonant-feminine (DF), and dissonant-masculine (DF).

A consonant object is, in real life, related to a gender that is consistent with the GG of its name in the critical language (i.e., Arabic). For instance, the object "skirt" is, in real life, more associated with women than men, and the word "skirt" is feminine in Arabic. Conversely, a dissonant object is, in real life, related to a gender that is opposite to the GG of its name. For instance, the object "tank" is, in real life, more associate with men than women, whereas the word "tank" is feminine in Arabic (B. Bassetti, 2007). Names of the pictures were verified by an independent group of 32 participants (25 native Arabic speakers and 7 native English speakers) and received an average of 95% < correct responses.

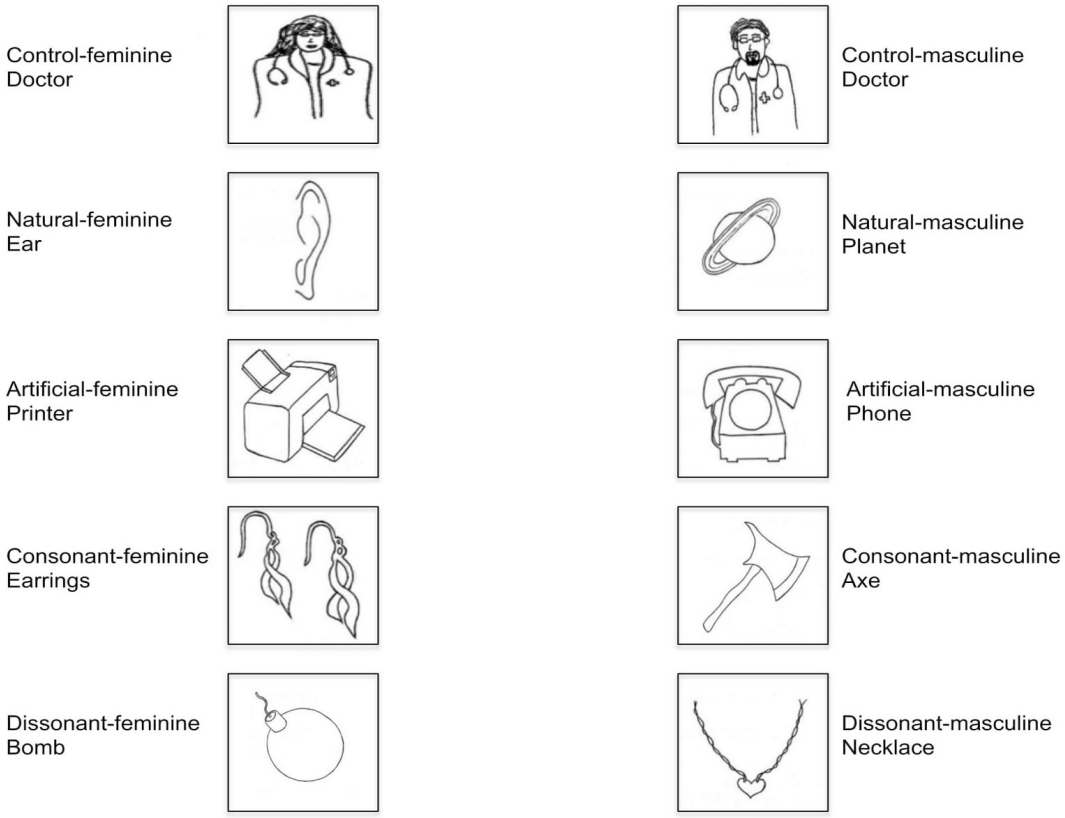

Figure 1

Examples of The Pictures Used in the Experiment.

The control objects were examples of pictures presented in the familiarisation phase, and the others were examples of pictures used in the experimental trials. The 'gender' (feminine and masculine) of the objects refers to the GG of their Arabic names rather than to their sex attributes, and the genders of the objects were balanced across all categories (natural, artificial, consonant, and dissonant). The natural and artificial categories were used to examine the natural-feminine/artificial-masculine hypothesis, while the consonant and dissonant categories were used to examine GG effects on object-gender stereotypical association.

Task and Procedure

A modified online version of the voice attribution task (VET), initially developed by Sera et al. (1994), and has been used previously (Kurinski et al., 2016; Sera et al., 2002; Vernich et al., 2017) was used in this experiment. The experiment was conducted using Qualtrics, an online survey software (https://www.qualtrics.com). After signing an online consent form, subjects were asked to give demographic information (i.e., age and gender), choose their native language between English and Arabic, and rate their level of proficiency in the second language before the experiment began. The

experiment began with the following instruction:

"Imagine that you are going to make a children's movie. In this movie, some everyday objects come to life. They will be able to talk, play and do everything. Imagine that you are the director of this movie. You will be presented with forty pictures of the inanimate objects that will be brought to life in the movie. Your task is to click on one of the two choices provided under each picture whether you think that it should have a man'svoice (or a boy's voice) or a woman's voice (or girl's voice)."

The instruction was given in Arabic for Arabic monolinguals and Arabic-English bilinguals. It was given in English for English monolinguals and English-Arabic bilinguals. After the instructions, the familiarisation phase began. Eight pictures were presented on the same page at once. Participants selected the voice that they would like to assign to the objects by clicking on their choice with the mouse. The test trials were administered in the same way as the familiarisation trials. There was no time limit for the experiment, and participants could adjust their choices during the experiment. Submission of the answers was not possible until every object had been assigned a voice. Participants were thanked for their participation when the experiment ended. The average duration of the experiment was approximately 7 minutes.

Results

The results of 2 native Arabic monolinguals, 2 native English monolinguals, and 1 English-Arabic bilingual with an advanced level of proficiency were excluded from analysis due to their poor performance in the familiarisation procedure as they failed to provide correct voice assignments to 75% of the control items, indicating that they did not understand the task (Sera et al., 1994; Sera et al., 2002). Responses from the remaining participants were then converted into scores (i.e., percentages by the Arabic GG system).

First, to investigate whether there would be a difference between native Arabic speakers and native English speakers on their performances in the voice attribution task, a three-factor analysis of variance (ANOVA) was conducted, with Language (Arabic, including Arabic monolinguals and Arabic-English bilinguals versus English monolinguals and English-Arabic bilinguals) as the between-subjects factor and category (natural, artificial, consonant, and dissonant) and grammatical gender (feminine, masculine) as the within-subjects factors. The result showed a main effect of language (VAT), $F(1,113)=49.360$, $p<0.0001$. Arabic speakers significantly attributed more voices to objects in consistence with Arabic GG ($M=67.03$, $SD=1.22$) compared to English speakers ($M=54.81$, $SD=1.23$). Additionally, there was a main effect of category on VAT performance ($F(3.339)=460.30$, $p<0.0001$). Pairwise comparisons showed that the consonant category was assigned significantly more voices ($M=93.88$, $SD=1.17$) consistently with Arabic GG compared to the natural ($M=66.22$, $SD=1.42$)($F(3,111)=272$, $p<0.0001$), artificial ($M=67.17$, $SD=1.37$) ($p<0.0001$), and dissonant categories ($M=16.39$, $SD=2.13$) ($p<0.0001$); in contrast, the dissonant category was significantly assigned the least number of voices compared to

other categories (ps<0.0001). However, natural and artificial categories did not significantly differ from one another (F (3,111)=272, p=0.630). There was also a main effect of grammatical gender (F (1,113)=11.263, p=0.001). A voice attribution to grammatically masculine objects was more consistent with Arabic GG (M=64.00, SD=1.12) compared to grammatically feminine objects (M=57.82, SD=1.40).

In addition, there was no significant interaction between grammatical gender and language (F (1,113)=0.025, p>0.05, but there was a significant interaction between category and language (F (3.339)=10.204, p<0.0001). Pairwise comparisons showed that Arabic speakers assigned more voices in accordance with Arabic GG compared to English speakers in the natural (p>0.0001), artificial (p>0.0001), consonant (P=0.006), and dissonant categories (p=0.272). However, while Arabic speakers assigned more voices to artificial objects (M=81.17, SD=1.93) compared to natural objects (M=71.98, SD=2.00) (F (3,111)=167.18, p=0.004), English speakers assigned more voices to natural objects (M=60.45, SD=2.02) compared to artificial ones (M=54.17, SD=1.95)(F (3.111)=130.052, p=0.026).

Additionally, there was a significant interaction between category and grammatical gender (F=(3,339)=45.103, p<0.0001). Pairwise comparisons showed that natural and artificial categories were assigned more voices for grammatically masculine objects compared to feminine objects (ps<0.0001). They also revealed that the consonant category was assigned slightly more feminine voices compared to masculine voices (p<0.05), while the dissonant category was assigned significantly more feminine voices compared to masculine ones (p<0.0001). Furthermore, voice attribution was higher for grammatically masculine natural objects (M=77.46, SD=2.04) compared to artificial ones (M=75.11, SD=2.11) (p>0.05) but was higher for grammatically feminine artificial objects (M=59.22, SD=2.50) compared to natural ones (M=54.98, SD=2.98) (p>0.05),

In addition, there was a significant interaction between category, grammatical gender, and language (F (3,339)=3.164, p=0.025). Arabic and English speakers showed a similar tendency to attribute more voices to grammatically masculine natural and artificial objects compared to feminine ones as well as to attribute more voices to grammatically feminine dissonant objects compared to masculine ones. However, Arabic speakers assigned more voices consistently with Arabic GG to artificial objects (both grammatically feminine and masculine) compared to natural ones, whereas English speakers assigned more voices to natural objects (both grammatically feminine and masculine) compared to artificial ones (see Table 2).

Table 2

Percentage of Times Objects Were Assigned Voices Consistent with Arabic Grammatical Gender by Arabic Speakers (Arabic Monolinguals And Arabic-English Bilinguals) and English Speakers (English Monolinguals And English-Arabic Bilinguals).

	Natural		Artificial		Consonant		Dissonant	
Language	F	M	F	M	F	M	F	M
Arabic Speakers	57.76	86.21	73.71	86.64	96.55	97.84	27.16	10.34
English Speakers	52.19	68.71	44.74	63.60	91.23	89.91	19.30	8.77

Labels used in the second row refer to feminine (F) and masculine (M).

Second, to investigate whether Arabic monolinguals and Arabic-English bilinguals would differ in performance on the VAT, another three-factor ANOVA was conducted with Language (Arabic monolinguals, Arabic-English bilinguals) as the between-subjects factor and category (natural, artificial, consonant, and dissonant) and grammatical gender (feminine, masculine) as the within-subjects factors. The result showed no main effect of language ($F(1,56)=1.352$, $p>0.05$). The VAT performance of Arabic monolinguals ($M=65.40$, $SD=1.94$) and Arabic-English bilinguals ($M=68.54$, $SD=1.88$) did not significantly differ from each other. Similar to the first analysis, there were main effects of category ($F(3,168)=327.79$, $p<0.000$) and grammatical gender ($F(1,56)=8.533$, $p=0.005$). Voice attribution was significantly higher for the consonant but lower for the dissonant category compared to other categories. Moreover, grammatically masculine objects were assigned significantly more voices compared to grammatically feminine ones.

Furthermore, there was a significant interaction between category and language ($F(3,168)=671.956$, $p<0.022$). Arabic-English bilinguals attributed more voices consistently with Arabic GG compared to Arabic monolinguals in all categories, except for consonant objects (see Table 3). However, while no significant interaction was found between grammatical gender and language ($F(1,56)=0.237$, $p>0.05$), a significant interaction was found between category and grammatical gender ($F(3,168)=39.56$, $p<0.0001$). Grammatically masculine objects of all categories were assigned more voices in accordance with Arabic GG compared to feminine, except for dissonant categories. There was no significant interaction between category, grammatical gender, and language ($F(3,168)=1.353$, $p>0.05$).

Table 3

Percentage of Times Objects Were Assigned Voices Consistent with Arabic Grammatical Gender by Arabic Monolinguals and Arabic-English Bilinguals

	Natural		Artificial		Consonant		Dissonant	
Language	F	M	F	M	F	M	F	M
Arabic Monolinguals	59.82	83.04	74.11	84.82	98.21	99.11	18.75	5.36
Arabic-English Bilinguals	55.83	89.17	73.33	88.33	95.00	96.67	35.00	15.00

Labels used in the second row refer to feminine (F) and masculine (M).

Third, to examine whether English monolinguals and English-Arabic bilinguals would differ in performance on the VAT, a three-factor ANOVA was conducted with Language (English monolinguals versus English-Arabic bilinguals) as the between-subjects factor and category and grammatical gender as the within-subjects factors. The result showed no main effect of language on the VAT performance ($F(1,55)=1.253$, $p>0.05$). The performance of English monolinguals ($M=53.57$, $SD=1.54$) and English-Arabic bilinguals ($M=56.00$, $SD=1.52$) did not significantly differ. Similar to previous findings, there were main effects of category ($F(3,165)=179.846$, $p<0.0001$) and grammatical gender ($F(1,55)=4.206$, $p=0.045$). The VAT performance of English monolinguals and English-Arabic bilinguals was significantly different based on category and was more consistent with Arabic for grammatically masculine objects compared to feminine ones.

Additionally, there was no significant interaction between category and language ($F(3.165)=0.388$, $p>0.05$), but a significant interaction was found between grammatical gender and language ($F(1,55)=8.110$, $p=0.006$). English monolinguals assigned more voices to grammatically feminine objects ($M=54.69$, $SD=2.55$) compared to masculine ones ($M=52.45$, $SD=2.50$), but English-Arabic bilinguals assigned more voices to grammatically masculine objects ($M=62.86$, $SD=2.46$) compared to feminine ones ($M=49.14$, $SD=2.50$), Additionally, there was a significant interaction between category and grammatical gender ($F(3,165)=14.163$, $p<0.0001$). Natural and artificial objects were assigned more masculine voices in accordance with Arabic GG compared to feminine ones, whereas consonant and dissonant objects were assigned more feminine voices compared to masculine ones. Additionally, there was a significant interaction between category, grammatical gender, and language ($F(3,165)=2.694$, $p=0.048$). English

monolinguals assigned voices to grammatically feminine objects of all categories at slightly higher rates than English-Arabic bilinguals did. However, English-Arabic bilinguals assigned voices to grammatically masculine objects of all categories at significantly higher rates than English monolinguals did ($p<0.05$), except for the consonant category (see Table 4).

Table 4

Percentage of Times Objects Were Assigned Voices Consistent with Arabic Grammatical Gender by English Monolinguals and English-Arabic Bilinguals

	Natural		Artificial		Consonant		Dissonant	
Language	F	M	F	M	F	M	F	M
English Monolinguals	55.36	61.61	51.79	55.36	91.96	90.18	19.64	2.68
English-Arabic Bilinguals	49.14	75.57	37.93	71.55	90.52	89.66	18.97	14.66

Labels used in the second row refer to feminine (F) and masculine (M).

Discussion and Conclusion

The current study examined the effect of a grammatical gender system on the conceptual representation of objects. An online version of the VAT prevents potential biases that may have arisen during interactions between the experimenter and participants given the finding that the sex of a speaker may influence the grammatical processing of listeners. Results showed that compared to native English speakers (including English-Arabic bilinguals), native Arabic speakers (including Arabic-English bilinguals) assigned voices of genders more consistently with the GG of the objects' Arabic names; the same pattern of findings was reported in previous studies that investigated different language groups (Forbes et al., 2008; Kurinski & Sera, 2011; Sera et al., 1994; Sera et al., 2002; Vernich et al., 2017). The between-groups difference suggests a generic effect of GG on object categorisation, supporting the linguistic relativity hypothesis.

The results also showed that Arabic GG effects varied based on category and can be classified into 3 category-related effects. First, the GG effect can guide performance in natural and artificial categories (Slobin, 1996). For instance, Arabic speakers assigned more voices of gender in accordance with Arabic GG to natural (71.88%) and artificial categories (80.17%) compared to English speakers (natural: 60.45%, artificial: 54.17%, respectively). The second type of GG effect is to enhance the object-gender stereotypical association when there is

an agreement between the conceptual and grammatical gender of objects (Vernich et al., 2017). Whereas both Arabic and English speakers were more likely to assign voices to the consonant category based on the object-gender stereotypical association, GG enhanced the performance of Arabic speakers, in the sense that they significantly attributed more voices to consonant objects (97.20%) compared to English speakers (90.57%).

Third, the GG effect can reduce the object-gender stereotypical association when there is a conflict between the conceptual and grammatical gender of object (Vernich et al., 2017). For instance, Arabic and English speakers assigned voices to the dissonant category based on the object-gender stereotypical association, but Arabic speakers assigned fewer voices to the dissonant category based on the object-gender stereotypical association (81.25%) compared to English speakers (85.97%). The performance of Arabic and English speakers on consonant and dissonant categories is consistent with previous studies using word connotation judgement tasks (Alexander Z Guiora, 1983; Alexander Z. Guiora & Sagi, 1978) and reaffirms that participants will assign the gender of voices to objects based on perceived connotative values of objects rather than on their GG.

The results also showed a major tendency to assign masculine voices to grammatically masculine objects (64.00%) compared to assigning feminine voices to grammatically feminine objects (57.83%) This tendency was also found and explained in terms of a possible non-arbitrary link between the Spanish grammatically masculine gender and the conceptual gender of objects (Kurinski & Sera, 2011; Sera et al., 1994). This explanation emerged in the context of evidence supporting the natural-feminine/artificial-masculine hypothesis (Mullen, 1990; Sera et al., 1994). However, the pattern of the current results is inconsistent with the natural-feminine/artificial-masculine hypothesis. Arabic and English speakers assigned masculine voices to grammatically masculine natural and artificial objects more than feminine voices to grammatically natural and artificial objects in contrast with the mixed findings of previous studies (e.g., (Forbes et al., 2008; Kurinski et al., 2015; Kurinski & Sera, 2011; Sera et al., 1994; Sera et al., 2002)).

The dominance of the masculine un-marked grammatical gender form may explain the masculine attribution bias. This explanation was used to explain the tendency of native Spanish speakers and Spanish language learners to assign more voices in accordance with Spanish GG compared to feminine ones. Similarly, the tendency of Arabic speakers and English-Arabic bilinguals to assign more masculine voices in accordance with Arabic GG may have resulted from the un-marked masculine gender, since no such tendency was found in the performance of English monolinguals. Although it would be possible to attribute differences in VAT performance between Arabic and English speakers to differences in their socially determined/preferred gender association of objects, the shared performance patterns on consonant and dissonant categories between Arabic and English speakers ruled out such a possibility since objects in the consonant and dissonant categories have gender-stereotypical associations that may be

more susceptible to social differences compared to the natural and artificial categories.

Additionally, similar to the findings of Lambelet (2016), the results showed that bilingualism did not interact with the GG effect, as the VAT performance of Arabic-English and English-Arabic bilinguals was highly comparable with that of Arabic and English monolinguals, respectively. This result differed from those of previous studies that showed an effect of GG in the second language on the VAT performance of bilinguals (Forbes et al., 2008; Kousta et al., 2008; Kurinski et al., 2015; Kurinski & Sera, 2011; Nicoladis & Foursha-Stevenson, 2011). One explanation for this discrepancy is related to participants' second-language proficiency (Kurinski et al., 2015; Kurinski & Sera, 2011). Only one-third of English-Arabic bilinguals rated their Arabic proficiency level as advanced, whereas the rest rated their Arabic proficiency as beginner or intermediate level. Such a low percentage of proficient English-Arabic bilinguals might have been the reason behind the weak effect of GG. The VAT performance of English speakers with an advanced proficiency level in Spanish was significantly more consistent with Spanish GG compared to that of English speakers with a beginner level of Spanish.

Another possible explanation of performance indifference between monolinguals and bilinguals is the language of task instructions. The performance of bilinguals was found to depend on task languages. Italian-English bilinguals' performance on an error-induction task was similar to that of English monolinguals when they were given the task in English and similar to Italian monolinguals when given the task in Italian. However, this explanation may not be suitable in this situation since the VAT performance of bilinguals was more consistent with Arabic GG compared to monolinguals, even though bilinguals and monolinguals were both given the task in same native language. One possible explanation is the metalinguistic awareness account put forward by Bassetti (2007): The knowledge of two language systems affects bilinguals' perception of reality. As compared to Arabic monolinguals, differences in the grammatical gender system between English and Arabic may have enhanced the awareness of a GG perspective in Arabic-English bilinguals when there is a conflict between the objects' conceptual and grammatical gender. In fact, the same trend can be observed in the performance of English-Arabic bilinguals, where the magnitude of this effect might have been reduced due to the proficiency in their second language.

Furthermore, the performance of English-Arabic bilinguals on the VAT suggests a grammatical gender-specific effect. Whereas the performance of English monolinguals and English-Arabic bilinguals did not significantly differ on the VAT, English-Arabic bilinguals significantly assigned voices to grammatically masculine objects in the natural, artificial, and dissonant categories in accordance with Arabic GG more than English monolinguals did. The English-Arabic bilinguals' tendency to assign more masculine voices in accordance with Arabic GG could be attributed to masculine gender dominance and its popular masculine unmarked form. A similar tendency of masculine attribution was found in the VAT performance of English-Spanish

bilinguals and Hungarian-Spanish bilinguals (Kurinski et al., 2016; Kurinski & Sera, 2011).

In a similar vein, the added markers of feminine gender may explain why English monolinguals and English-Arabic bilinguals did not differ in assigning voices to grammatically feminine objects in all categories. While the GG effect depends on language proficiency (Forbes et al., 2008; Kurinski et al., 2016), learners of gendered languages (e.g. Spanish) have been shown to be less accurate in dealing with feminine nouns compared to masculine ones (Montrul, Foote, & Perpiñán, 2008) and to assign fewer feminine voices consistently with GG compared to masculine voices (Kurinski & Sera, 2011). Interestingly, language proficiency level seems to play a less important role when speakers deal with grammatically masculine nouns, based on the results of Seigneuric, Zagar, Meunier, & Spinelli (2007), which showed a tendency to assign masculine voices to unfamiliar nouns.

In summary, the current study provides a demonstration of GG effects on object categorisation in native Arabic speakers, corroborating previous findings of linguistic relativity studies on Indo-European languages. However, Arabic and English speakers showed a tendency to associate male voices with both natural and artificial objects regardless of the GG in Arabic. This pattern of results is inconsistent with the natural-feminine/artificial-masculine hypothesis, indicating a masculine grammatical gender bias of Arabic GG. Linguistic knowledge (i.e., GG) of a second language did not affect object categorisation either in a gender-positive direction (English-Arabic bilinguals) or a gender-negative direction (Arabic-English bilinguals). In both groups of Arabic speakers, object categorisation is primarily determined by the associative stereotypical gender of objects in case it conflicts with the GG. Taken together, on one hand, our findings argue against the language encapsulation view (Chomsky, 2000; J. Fodor, 1975; J. A. Fodor, 2008; Pinker, 1995, 2007), but, on the other hand, social factors (e.g., object-gender stereotypical association) may constrain and override the effect of GG. Future studies will need to further specify interactions between conceptual variables and linguistic structure on supporting people's cognitive organisation. Future studies also should address the current study's limitations, which include the small number of objects per category and self-evaluation of proficiency in second language.

Declaration of Interest

The authors declare no conflict of interests.

Acknowledgement

This is a research project that was supported by a grant from the research center for college of education, deanship of scientific research at King Saud University.

Reference

Al-Rasheed, A., Franklin, A., Drivonikou, G., & Davies, I. (2014). Left hemisphere lateralization of categorical color perception among roman and Arabic script readers. *Psychology,* 2014.

Alkohlani, F. A. (2016). The problematic issue of grammatical gender in Arabic as a foreign

language. *Journal of Language and Cultural Education, 4*(1), 17-28.

Andonova, E., Gosheva, A., Janyan, A., & Schaffai, J. S. (2007). Second language gender system affects first language gender classification. In *Cognitive aspects of bilingualism* (pp. 271-299): Springer.

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and cognition, 9*(1), 89-96.

Bassetti, B. (2007). Bilingualism and thought: Grammatical gender and concepts of objects in Italian-German bilingual children. *International Journal of Bilingualism, 11*(3), 251-273.

Bassetti, B. A. (2014). Is grammatical gender considered arbitrary or semantically motivated? Evidence from young adult monolinguals, second language learners, and early bilinguals. *British Journal of Psychology, 105*(2), 273-294.

Beit-Hallahmi, B., Catford, J., Cooley, R. E., Dull, C. Y., Guiora, A. Z., & Raluszny, M. (1974). Grammatical gender and gender identity development: cross cultural and cross lingual implications. *American Journal of Orthopsychiatry, 44*(3), 424.

Brysbaert, M., Fias, W., & Noel, M.-P. (1998). The Whorfian hypothesis and numerical cognition: istwenty-four'processed in the same way asfour-and-twenty'? *Cognition, 66*(1), 51-77.

Chomsky, N. (2000). *New horizons in the study of language and mind*: Cambridge University Press.

Clarke, M. A., Losoff, A., McCracken, M. D., & Still, J. (1981). Gender perception in Arabic and English. *Language Learning, 31*(1), 159-169.

Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The effect of grammatical gender on object categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 449.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398*(6724), 203-204.

Davies, I. R. (1998). A study of colour grouping in three languages: A test of the linguistic relativity hypothesis. *British Journal of Psychology, 89*(3), 433-452.

Ellis, L., & Ficek, C. (2001). Color preferences according to gender and sexual orientation. *Personality and Individual Differences, 31*(8), 1375-1379.

Ervin, S. M. (1962). The connotations of gender. *Word, 18*(1-3), 249-261.

Flecken, M. (2011). Event conceptualization by early Dutch–German bilinguals: insights from linguistic and eye-tracking data. *Bilingualism: Language and cognition, 14*(01), 61-77.

Fodor, J. (1975). The Language of Thought (Cambridge, MA: Harvard). In: University Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*: Oxford University Press on Demand.

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*, 68-76.

Garvin, P. L. (1958). LINGUISTICS: Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. John B. Carroll (Ed.). *American Anthropologist, 60*(2), 415-416. doi:10.1525/aa.1958.60.2.02a00480

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496-499.

Guiora, A. Z. (1983). Language and concept formation: A cross-lingual analysis. *Cross-Cultural Research, 18*(3), 228-256.

Guiora, A. Z., & Sagi, A. (1978). A CROSS-CULTURAL STUDY OF SYMBOLIC MEANING–DEVELOPMENTAL ASPECTS1. *Language Learning, 28*(2), 381-386. doi:10.1111/j.1467-1770.1978.tb00141.x

He, H., Li, J., Xiao, Q., Jiang, S., Yang, Y., & Zhi, S. (2019). Language and perception: Evidence from Mongolian and Chinese speakers. *Frontiers in psychology, 10*, 551.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of psycholinguistic research, 22*(5), 519-534.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 843.

Kurinski, E., Jambor, E., & Sera, M. D. (2015). Spanish grammatical gender: Its effects on categorization in native Hungarian speakers. *International Journal of Bilingualism*, 1367006915576833.

Kurinski, E., Jambor, E., & Sera, M. D. (2016). Spanish grammatical gender: Its effects on categorization in native Hungarian speakers. *International Journal of Bilingualism, 20*(1), 76-93.

Kurinski, E., & Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism: Language and cognition, 14*(02), 203-220.

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

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in cognitive sciences, 8*(3), 108-114.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-

naming systems. *Psychological Science, 6*(1), 56-60.

Moeller, K., Shaki, S., Göbel, S. M., & Nuerk, H.-C. (2015). Language influences number processing–a quadrilingual study. *Cognition, 136*, 150-155.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3), 503-553.

Mullen, M. K. (1990). Children's classifications of nature and artifact pictures into female and male categories. *Sex Roles, 23*(9-10), 577-587.

Nicoladis, E., & Foursha-Stevenson, C. (2011). Language and culture effects on gender classification of objects. *Journal of Cross-Cultural Psychology*, 0022022111420144.

Papafragou, A., & Selimis, S. (2010). Event categorisation and language: A cross-linguistic study of motion. *Language and cognitive processes, 25*(2), 224-260.

Picariello, M. L., Greenberg, D. N., & Pillemer, D. B. (1990). Children's Sex-related Stereotyping of Colors. *Child Development, 61*(5), 1453-1460.

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529): Penguin UK.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*: Penguin.

Sera, M. D., Berge, C. A., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development, 9*(3), 261-292.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General, 131*(3), 377.

Slobin, D. I. (1996). From" thought and language" to" thinking for speaking.".

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567-4570.

Vernich, L., Argus, R., & Kamandulytė-Merfeldienė, 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.

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: General, 134*(4), 501.

Vitevitch, M. S., Sereno, J., Jongman, A., & Goldstein, R. (2013). Speaker sex influences processing of grammatical gender. *PloS one, 8*(11), e79701.

Xia, T., Xu, G., & Mo, L. (2019). Bi-lateralized Whorfian effect in color perception: Evidence from Chinese Sign Language. *Journal of Neurolinguistics, 49*, 189-201.

Özgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science, 13*(3), 95-98.

Özgen, E., & Davies, I. R. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General, 131*(4), 477.