Grammatical gender affects gender perception: Evidence for the structural-feedback hypothesis

Sayaka Sato & Panos Athanasopoulos

Lancaster University, Department of Linguistics and English Language

Bailrigg, Lancaster, LA1 4YL, U.K.

s.sato1@lancaster.ac.uk (Sayaka Sato)

p.athanasopoulos@lancaster.ac.uk (Panos Athanasopoulos)

Author Note

Correspondence concerning this article should be addressed to Sayaka Sato, Lancaster University, Department of Linguistics and English Language, Bailrigg, Lancaster, LA1 4YL, U.K. E-mail: s.sato1@lancaster.ac.uk

Abstract

Two experiments assessed the extent to which grammatical gender provides a predictive basis for bilinguals' judgments about perceptual gender. In both experiments, French-English bilinguals and native English monolinguals were consecutively presented with images of objects manipulated for their (i) conceptual gender association and (ii) grammatical gender category and were instructed to make a decision on a subsequent target face. The experiments differed in the implicitness of the association between the object primes and target faces. Results revealed that when prior knowledge sources such as conceptual gender can be strategically used to resolve the immediate task (Experiment 1), this information was readily extracted and employed. However, grammatical gender demonstrated a more robust and persisting effect on the bilinguals' judgments, indicating that the retrieval of obligatory grammatical information is automatic and modulates perceptual judgments (Experiment 2). These results suggest that grammar enables an effective and robust means to access prior knowledge which may be independent of task requirements.

Keywords: linguistic relativity; bilingualism; categorization; grammatical gender; conceptual gender

Grammatical gender affects gender perception: Evidence for the structural-feedback hypothesis

1. Introduction

The notion that the languages we speak are responsible for shaping our thoughts can be traced back to Whorf's classic principle of linguistic relativity (Whorf, 1956). A view that has stirred heated discussions regarding the extent of its influence (Pinker, 1994), recent questions are now geared toward understanding how languages may contribute in modulating non-verbal cognition (Athanasopoulos, Bylund, & Casasanto, 2016). Specifically, language or labels (i.e., words) are now characterized as offering a conceptual basis that motivates the top-down processing of perceptual information (e.g., Lupyan, 2012; Lupyan & Clark, 2015). Although a substantial body of studies on the representation of color (e.g., Davidoff, Davies, & Roberson, 1999; Roberson, Davies, & Davidoff, 2000; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; Winawer et al., 2007), time (e.g., Boroditsky, 2000; Casasanto & Boroditsky, 2008) and number (e.g., Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Frank, Everett, Fedorenko, & Gibson, 2008) point in favor of such a view, existing research has not yet been able to fully characterize the scope in which more complex linguistic features such as grammar may permeate and guide our cognitive processes.

In the study reported here, we sought to assess the extent to which grammatical information influences perceptual judgments by employing grammatical gender and conceptual gender information as testbeds to guide our investigation. Grammatical gender refers to a system of assigning noun class found in a vast majority of the languages in the world (Corbett, 1991). Contrary to languages such as English which do not incorporate such grammatical systems, grammatical gender languages such as French arbitrarily assign all nouns to a formal grammatical category (e.g., grammatically masculine: *couteau* [knife] vs. grammatically feminine: *cuillère* [spoon]). In contrast, conceptual gender concerns the conceptual properties of an object relating to either gender (e.g., conceptually male: *hammer* vs. conceptually female: *necklace*) which is not determined by linguistic or natural (i.e., biological) gender categories (Sera, Berge, & del Castillo-Pintado, 1994). Given that gender information spans on both grammatical and conceptual levels of representation, it provides a convenient case to evaluate the relationship between language and thought.

In fact, grammatical gender has been commonly employed to fuel the debate on linguistic relativity (see Cubelli, Paolieri, Lotto, & Job, 2011 for an exhaustive review of different empirical paradigms). Studies employing voice attribution (e.g., Flaherty, 2001; Sera et al., 1994), trait attribution (e.g., Boroditsky, Schmidt, & Phillips, 2003; Konishi, 1993) and inference generation (e.g., Imai, Schalk, Saalbach, & Okada, 2014; Saalbach, Imai, & Schalk, 2012) tasks suggest that language users rely on grammatical gender membership of an entity to infer its sex-related properties, even in cases where gender information should not be relevant. Although these studies provide evidence to suggest that grammatical information is readily mapped onto an entity's semantic representation, the implemented paradigms are explicit and constrain these findings to cases where speakers consciously engage in verbal processing.

This is not to say, however, that grammatical features exhibit only superficial effects on general cognition. Theoretical frameworks such as connectionist approaches provide an alternative explanation to the contribution of language on thought. Under this notion, the human mind is characterized as being highly interactive, where features such as language are considered to play an active role during the encoding of concepts and categories (e.g., Rumelhart, McClelland, & PDP Research Group, 1986). Lupyan's (2012) label-feedback hypothesis draws upon such notions, particularly emphasizing the role of language, suggesting that categorical labels (i.e., words) modulate visual perception. Learning to associate properties of an entity with a specific label allows the perceiver to abstract distinctive features of a given exemplar to a more typical category. Predictions triggered from the label are thus activated and successively fed back in a top-down manner, causing a temporary modulation on on-line perceptual representations. Hearing a redundant label in a visual search task, for example, has been shown to enhance detection by directing attentional focus to the stimuli's prototypical features (e.g., Lupyan & Spivey, 2010a, 2010b; Lupyan & Ward, 2013). Labels are thus expected to streamline perceptual representation by heightening features of the relevant stimuli.

Notwithstanding the broad appeal of Lupyan's framework, it does not fully account for the structural influences brought on by linguistic features such as grammar and syntax, which are essentially more complex than single labels (Thierry, 2016). These considerations are critical; language processing is inherently dynamic, inasmuch as locally attending to structural and agreement relationships is requisite and obligatory for the language user (Lucy, 1997). To the extent that the exposure to languages or acquisition of new linguistic constructions may even restructure our conceptual representations (e.g., Athanasopoulos, 2006; Majid, Bowerman, Kita, Haun, & Levinson, 2004), it is unlikely that the effects of language are restricted to the retrieval of labels.

Indeed, a wealth of recent studies has shown that perceptual modulations can bear the consequence of syntactic and grammatical encodings of specific languages (e.g., Athanasopoulos & Bylund, 2013; Boutonnet, Athanasopoulos, & Thierry, 2012; Casasanto & Boroditsky, 2008; Fausey & Boroditsky, 2010). This is exemplified by research on motion event construal demonstrating a link between perceivers' attentional predispositions and the syntactic framing of motion events in their respective languages (e.g., Athanasopoulos et al., 2015; Athanasopoulos & Bylund, 2013; Flecken, Athanasopoulos, Kuipers, & Thierry, 2015). For example, Flecken et al. (2015) compared attention allocation during motion event perception among English and German speakers. Given that English linguistically emphasizes trajectory and endpoint of motion as opposed to only endpoints in German, their study found distinct language-consistent preferential biases of motion aspect. Specifically, attention was more heavily allocated to aspects that were linguistically encoded in each language, although the task did not require any conscious verbal processing. More relevant to the present study is evidence provided by Boutonnet et al. (2012), who demonstrated that morphological properties such as grammatical gender impose a significant impact on categorization. In their study, native English speakers and Spanish-English bilinguals were presented with three object images and were instructed to judge if the third target object image belonged to the same semantic category to the previous two objects. They found that while all participants were sensitive to the semantic associations between the objects, the bilinguals were also affected by the hidden manipulation of their grammatical gender membership in Spanish. The authors concluded that language-based properties were automatically accessed during object categorization and were subsequently fed back into lower-level perceptual processes even in conditions where linguistic mediation was unwarranted.

The aforementioned studies provide substantial evidence pointing to the inherent complexity of the top-down influences of language. Flecken et al. (2015) acknowledged the possibility that the different perceptual biases observed in their study could have originated from speakers essentially labeling event trajectory and endpoints. This would allow label-feedback effects to arise (i.e., labels activating diagnostic features) and preserve the reported language-specific properties. However, the findings of Boutonnet et al. (2012) confirm that grammatical gender is covertly recruited, and suggest that the information that is fed back to the perceptual system extends far beyond the impact that mere labels may activate. Assuming that grammatical gender categories operate as an obligatory and formal grammatical cue, it stands to reason that they may cast a more significant influence on perceivers' categorical decisions than would single labels.

The effects of language can thus be characterized as predisposing perceivers' attention to aspects that are linguistically realized, with grammar providing a *structural feedback* that guides our perceptual processes. In other words, although grammatical gender may lack semantic relevance to the lexical or conceptual representation of an entity, it nonetheless is a salient and obligatory feature. The encoding of this information and the constant attention it warrants would most likely structure a speakers' tendency in adhering to this category. Consequently, this would result in guiding or biasing the perceptual categorization of incoming information. Such a view ties in well with a predictive processing account that considers prior knowledge as modulating our perceptual representations (Lupyan & Clark, 2015). According to this framework, prior knowledge is rapidly recruited from longterm memory, allowing perceivers to generate probable expectations about incoming sensory information. In doing so, the information is contextualized through meaningful predictions, moderating predictive errors that are otherwise expected to arise. An intricate interaction is therefore expected between top-down predictions and incoming information, allowing perceivers to refine their perceptual experiences. For instance, our prior knowledge that dogs bark enables us to discern dogs from other furry animals that may resemble a similar entity. By the same token, language and their grammatical structures should activate information relevant to a given situation, affording more efficient means to retrieve top-down predictions.

The study reported here focuses on understanding the extent to which such grammatical structures may penetrate the biases emerging in our perceptual judgments. Specifically, we compared the influences grammatical gender may impose on categorization to that of non-linguistic prior knowledge about an object's associated conceptual gender. Past studies have shown that prior knowledge about a prime's conceptual gender information influences judgments in assessing the gender typicality of target facial stimuli, as demonstrated by exposure to gendered objects (Utz & Carbon, 2015), hormones (Kovács et al., 2004), and speech frequency (E. L. Smith, Grabowecky, & Suzuki, 2007). Here, two experiments utilizing an object triad task were carried out in an English-exclusive environment, to determine the extent to which grammatical gender would modulate perceptual judgments about the sex of facial stimuli. French-English bilinguals were compared against monolingual English-speaking controls where only English was spoken. Because English is not marked for grammatical gender, any grammatical gender effects that may be observed among the French-English bilinguals would provide evidence of the bilinguals' usage of language that is not being actively employed. In this manner, we aimed to provide evidence as to how grammatical gender may modulate cognitive processes during a task that did not necessitate its activation. In each experiment, participants were primed with two object images strongly associated with a conceptual gender and were instructed to make a sex-related judgment on a target facial image. The objects were manipulated for their (i) conceptual gender association and (ii) grammatical gender category.

In Experiment 1, participants were required to link the conceptual gender association of the object primes to the sex of the subsequent target face, judging whether the objects made them think of the target face or not. For both native English and French-English bilinguals, we predicted that prior knowledge about conceptual gender associations of objects would prime sex-related judgments on the target face. Critically, if grammatical gender information of the object is similarly spontaneously recruited, French-English bilinguals' judgments should also be modulated by its grammatical gender information. These effects were expected to emerge more robustly in cases where participants were required to strategically use gender information to resolve the task.

If indeed, grammatical gender is covertly recruited to modulate the French-English bilinguals' perceptual representations, the strength of this penetrability onto non-verbal tasks would need to be assessed. Thus, Experiment 2 attenuated the perceptibility of the gender association between the prime and target, such that the participants simply judged whether a target genderless face appeared to have more female or male-like traits while simply being exposed to object gender information. If the retrieval of grammatical information occurs unconsciously and permeates to perceptual processes irrespective of the preceding object primes, information that is activated may potentially and exclusively modulate the bilinguals' perceptual judgments on an unrelated task. Although both grammatical and conceptual gender were assumed to motivate top-down predictions on participants' judgments, the former was nonetheless expected to show an enhanced influence for French-English bilinguals, given that French incorporates, but English lacks grammatical gender.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty native English speakers (19 women; M_{age} = 22, age range = 18 – 47 years) who did not speak any other languages and 28 French-English bilinguals (14 women; M_{age} = 22, age range = 18 – 32 years) from Lancaster University (U.K.) took part in the study for monetary compensation. In order to assess second language (L2) proficiency levels, the French-English bilinguals completed a language background questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) as well as the Oxford Quick Placement Test (QPT: Oxford University Press, 2001), a standardized English proficiency test. Mean scores on the QPT reached 77.24% (*SD*: 12.23) which is equivalent to a C1 or advanced level on the Common European Framework of Reference for Languages (CEFR). Self-assessment of the bilinguals' L2 proficiency in speaking, reading and understanding of spoken language was rated as being adequate or above. All participants gave their informed written consent. Information regarding the bilinguals' language profile is summarized in Table 1. All participants had normal or corrected to normal vision. For the analyses, one participant was excluded from each language group because they did not understand the task instructions.

	Experiment 1	*Experiment 2*
Oxford Quick Placement Test score (%)	77.24 (12.23)	76.81 (14.79)
Age of acquisition of L2 English (yrs.)	8.57 (3.82)	9.91 (2.67)
Exposure to L2 English (%)	63.73 (15.77)	62.52 (17.99)

Note: Values are indicated as means (*M*) with their standard deviation (*SD*) in parentheses

Table 1. *Language profile of the French-English bilinguals in Experiments 1 and 2*

2.1.2. Materials and Design

The experimental task consisted of a triad of images, in which two object images were successively presented as gender primes, followed by a third target facial image. Having pairs of object images as opposed to one image not only served to obscure the hidden grammatical gender manipulation as suggested by Boutonnet et al. (2012) but allowed us to generate gender primes that had similar conceptual gender strength across all items. Thus, 240 object images judged to have a female (e.g., *ring*), male (e.g., *cigar*), or neutral (e.g., *spoon*) conceptual gender association were initially chosen from the Bank of Standardized Stimuli (BOSS; Brodeur et al., 2012; Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010). As for the target faces, four typically male and four typically female normed faces were obtained from Sato, Gygax, and Gabriel (2016).

All pictures were presented in greyscale and on a white backdrop to eliminate color biases. To assess the conceptual gender associations and the strengths of each object image, 13 native English and 10 native French speakers who did not take part in the main experimental session evaluated each object image on a 7-point Likert scale ranging from "very feminine" (1) to "very masculine" (7). Based on this pretest, items that were judged as having similar gender association strength in each language were selected. This resulted in 32 neutral (*M* = 3.98, *SD* = .89), 32 prototypically female (*M* = 2.89, *SD* =1.27), and 32 prototypically male (*M* = 5.38, *SD* = 1.23) items, of which half were grammatically masculine and the other half grammatically feminine in French in each conceptual gender condition (e.g., neutral conceptual gender: *spoon* [*cuillère* grammatically feminine] vs. *soap* [*savon* grammatically masculine]; female conceptual gender: *pushchair* [*poussette* grammatically feminine] vs. *necklace* [*collier* grammatically masculine]; male conceptual gender: *necktie* [*cravate* grammatically feminine] vs. *hammer* [*marteau* grammatically masculine]).

Objects with the same conceptual gender were subsequently paired together to construct 16 object primes for each conceptual gender. The pairings took into consideration the rating of the conceptual gender association allocated to each object such that objects with strong associations were paired with those with weaker associations. This ensured that the prime pairs had relatively similar strengths in gender association. The gender association of the three conceptual gender conditions differed significantly in their mean ratings (neutral primes: *M* = 4, *SD* = .06; female primes: *M* = 2.23, *SD* = .28; male primes: *M* = 5.77, *SD* = .08, all *ps* < .001). Crucially, half of the object primes in each conceptual gender condition consisted of pairs that were grammatically feminine and the other half grammatically masculine in the French language. None of the pairs consisted of individual objects with different grammatical gender class. Finally, each object prime was combined with a male and a female facial image to construct two separate trials. The first two objects were also inverted for their order to ensure all possible combinations, producing 192 experimental triads. The fact that each object prime was paired with both a male and female face had the advantage that the object primes could be allocated to both conceptually gender-related or unrelated conditions as well as a grammatically gender congruent or incongruent condition. In doing so, the added variance caused by individual object primes were better controlled.

This yielded the following experimental conditions as detailed in Table 2 (see also Appendix A: Table A.1 for all experimental items per condition) where the object prime was: (1) conceptually gender-neutral to the target face sex and belonged to a (1a) grammatically gender congruent category or (1b) grammatically gender incongruent category to that of the target face sex, (2) conceptually gender-related to the target face sex and belonged to a (2a) grammatically gender congruent category or (2b) grammatically gender incongruent category to that of the target face sex, or (3) conceptually gender-unrelated to the target face sex and belonged to a (3a) grammatically gender congruent category or (3b) grammatically gender incongruent category to that of the target face sex. Finally, 18 additional filler items composed of two objects with either of the three possible conceptual genders with masculine or feminine grammatical gender were generated. For these items, a prompt asking whether the primes were human-made ("*Human-made*?") was presented following the object images instead of a facial image. As was the case for the experimental items, the order of the first two objects was inverted, resulting in a total of 36 filler items. These items were intended to motivate participants to process each stimulus properly.

Experimental Conditions		Possible Item Combinations		
Conceptual gender relatedness	Grammatical gender congruency	Conceptual gender of object prime	Grammatical gender of object prime	Target face sex
(1) *Neutral*	(1a) *Congruent*	Neutral	Feminine	Female
		Neutral	Masculine	Male
	(1b) *Incongruent*	Neutral	Masculine	Female
		Neutral	Feminine	Male
(2) *Related*	(2a) *Congruent*	Female	Feminine	Female
		Male	Masculine	Male
	(2b) *Incongruent*	Female	Masculine	Female
		Male	Feminine	Male
(3) *Unrelated*	(3a) *Congruent*	Female	Masculine	Male
		Male	Feminine	Female
	(3b) *Incongruent*	Female	Feminine	Male
		Male	Masculine	Female

Table 2. *Experimental conditions and their corresponding gender combination of items in Experiment 1*

2.1.3. Procedure

The stimuli were mounted on an Apple MacBook Pro running SR Research Experiment Builder (Version 1.10.1630) with responses recorded by a key press marked "*yes*" or "*no.*" Participants' dominant hand always corresponded with the "*yes*" key. Each trial began with a 1000 ms fixation cross, succeeded immediately by the presentation of the first object lasting 1000 ms, followed by a blank screen of 250 ms, and the second object lasting 1000 ms. Finally, a blank screen of 350 ms was presented before the final facial image appeared, which remained on the screen until a response was elicited or 4000 ms elapsed (see Figure 1 for the procedural timeline and example stimuli). Participants were instructed as follows: "Please decide as quickly and as accurately as possible whether the two object images make you think of the person represented by the facial image." The same presentation conditions were applied for the filler trials, and participants judged whether the two objects were human-made or not after the presentation of the prompt. All items were randomized for each participant.

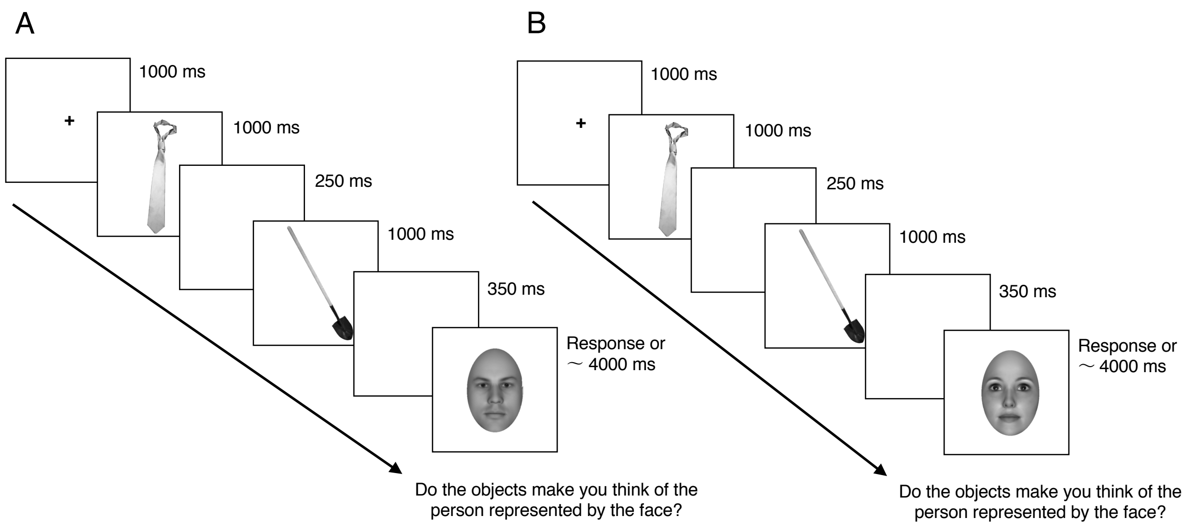


Figure 1. Experimental timeline for critical trials in Experiment 1. After the presentation of a fixation point (1000 ms), the first object was presented for 1000 ms, followed by a blank screen of 250 ms and the second object appeared for 1000 ms. After a blank screen of 350 ms following the second object, a male or female facial image was presented. Participants judged whether the objects made them think of the person represented by the face with a *yes* or *no* key press. The facial image remained on the screen until a response was initiated or 4000 ms elapsed. All object pairs were paired each with (A) a male and (B) a female face.

Participants were tested individually in a quiet room with all instructions given in English to ensure that the bilingual group did not associate the task with their first language. Prior to the main experimental task, participants were given four practice trials to familiarize themselves with the procedures of the experiment. The main task was followed by the completion of the LEAP-Q (Marian et al., 2007) and QPT (Oxford University Press, 2001) for the French-English bilinguals. The experimental session lasted approximately 45 minutes to one hour including debriefing.

2.2. Results

Given that participants' responses in the experiment depended heavily on their subjective opinions, all response times to target facial images were analyzed (see Figure 2 for mean response proportions for each level of conceptual gender relatedness). The decision to include all response times follows the analyses conducted by Boutonnet et al. (2012) but is particularly relevant within the context of this study. This is based on the assumption that grammatical gender is, in many cases, orthogonal to conceptual gender due to its arbitrariness. It should have therefore been activated both when participants conformed to or violated an expected conceptual gender association, and not only when they cohered with the expected gender association.

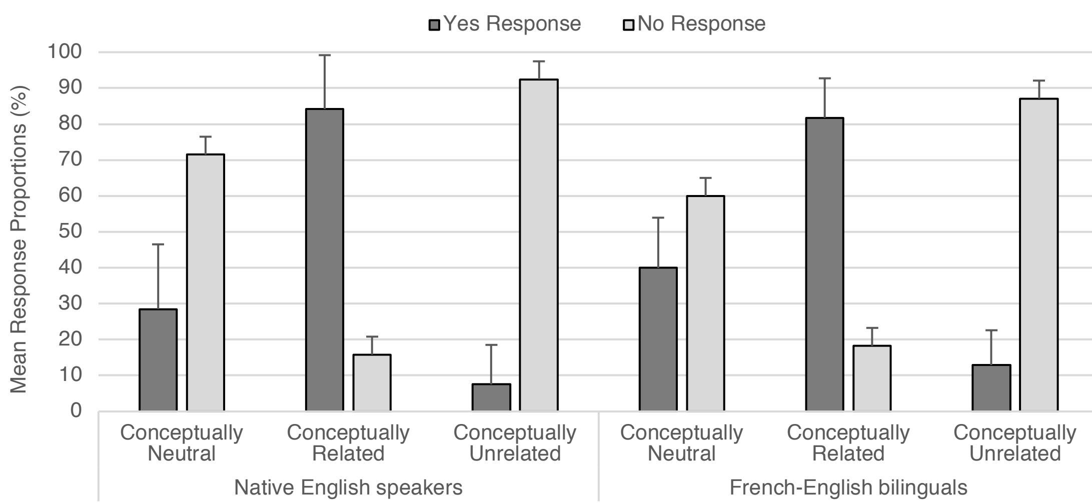


Figure 2. Proportion of response choices to target facial image in Experiment 1 for each language group. Error bars indicate standard error of the mean.

Response time data exceeding 3 *SDs* away from each participants' mean and shorter than 200 ms were excluded from the analyses. This resulted in the removal of 1.36 % of the native English speakers' data and 2.28 % of the French-English bilinguals' data. Mean response times in each condition for both language groups are shown in Figure 3.

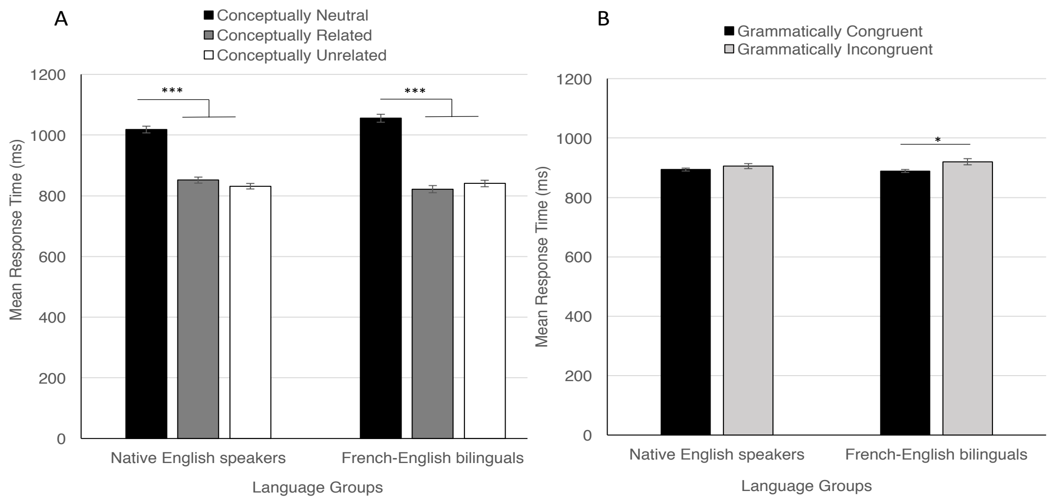


Figure 3. Mean response times (ms) of Experiment 1 in the (A) conceptual gender relatedness and (B) grammatical gender congruency condition. Error bars indicate standard error of the mean. $*p <.05$. $**p < .01$. $***p < .001$.

Linear mixed-effects models were subsequently fitted on the data separately for each language group (Baayen, Davidson, & Bates, 2008). Applying mixed-effects models allows to account for both by-participants and by-items variance simultaneously and are better at dealing with missing data than traditional ANOVA analyses. All analyses were conducted in the R environment (R Core Team, 2013) with the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) and the significance of fixed effects was calculated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016). As suggested by Barr, Levy, Scheepers, and Tily (2013), the models consisted of a maximal random effect structure with random intercepts and slopes for participants and items for the manipulated experimental predictors. If the model failed to converge, predictors were removed from the random structure successively.

Specifically, the analyses assessed the effects of *conceptual gender relatedness* of the objects to the target face sex (3: neutral / related / unrelated) and the *grammatical gender congruency* of the objects with the target face sex (2: congruent / incongruent) and their interactions. Given the three levels, conceptual gender relatedness was coded with orthogonal contrasts. Specifically, as participants were expected to have a firm idea of the conceptual gender relatedness of the objects, the decision to respond either yes or no was expected to indicate an equal behavioral index to the gender association. Consequently, both conceptually gender-related and gender-unrelated triads were expected to be responded faster than gender-neutral triads. Thus, the first contrast compared gender-laden (average of gender-related and unrelated) triads to that of gender-neutral triads and the second contrast compared the effect of gender-unrelated triads to that of gender-related triads. Grammatical gender congruency was sum coded (congruent: 0.5, incongruent: -0.5). The fixed effects structure of the final model for each language group is summarized in Table 3.

	Native English speakers				*French-English bilinguals*			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
(Intercept)	889.36	48.76	18.24	< .001	902.18	60.09	15.01	< .001
CG relatedness (L vs. N)	- 175.59	26.39	- 6.65	<.001	- 219.23	15.42	-14.22	< .001
CG relatedness (R vs. U)	26.14	27.54	0.95	*ns*.	- 17.42	14.50	-1.2	*ns*.
GG congruency (C vs. I)	- 6.87	9.74	- 0.71	*ns*.	- 31.73	15.36	- 2.07	< .05
CG relatedness (L vs. N) : GG congruency (C vs. I)	13.71	18.48	0.74	*ns*.	24.66	25.29	0.98	*ns*.
CG relatedness (R vs. U) : GG congruency (C vs. I)	- 34.19	24.14	- 1.42	*ns*.	43.36	35.43	1.22	*ns*.

Note: CG = Conceptual gender; GG = Grammatical gender; L = Gender-laden; N = Gender-neutral; R = Gender-related; U = Gender-unrelated; I = Grammatically Incongruent; C = Grammatically Congruent

Table 3. *Fixed effects estimates of mixed-effects model of response times for each language group (Experiment 1)*

2.2.1. Native English speakers

The final model for the native English speakers included by-participant and by-item intercepts, as well as by-participant random slopes for conceptual gender relatedness and grammatical gender congruency. Consistent with our hypothesis, the analyses revealed that responses were significantly slower for conceptually gender-neutral than gender-laden triads ($p < .001$), with no differences seen between gender-related and gender-unrelated triads (*ns*.). Triads with gender-specific information facilitated participants' judgments compared to gender-unspecific triads. This was to be expected, as participants would have activated a specific gender immediately after being presented with the object primes and needed to simply accept or reject the gender association for gender-laden triads. This, however, was not the case for gender-neutral triads which did not match or mismatch their expectations. Critically, the effect of grammatical gender congruency did not affect the response times (*ns.*), nor did it interact with conceptual gender relatedness (*ns.*). These effects were consistent with our predictions, as English does not mark for grammatical gender.

2.2.2. French-English bilinguals

The final model for the French-English bilinguals included by-participant and by-item intercepts, as well as by-participant and by-item random slopes for grammatical gender congruency. As was the case for the native English speakers, the analyses revealed that conceptually gender-neutral triads took significantly longer time to judge than gender-laden triads ($p < .001$). There were also no differences between gender-related and gender-unrelated triads (*ns.*). Crucially for the study, a significant effect of grammatical gender congruency emerged ($p < .01$), indicating that responses to triads with congruent grammatical gender were responded faster than incongruent triads irrespective of their conceptual gender congruency (interaction *ns*.).

2.3. Discussion

The categorization task presented in Experiment 1 required participants to make explicit decisions about the relationship between the object primes and the target face. We found that both native English speakers and French-English bilinguals extracted conceptual gender information of the objects which significantly impacted the judgments on target facial images. Importantly, the effect of grammatical gender congruency emerged exclusively for the French-English bilinguals, confirming the retrieval of grammatical information when making non-verbal categorical judgments. These results corroborate the findings by Boutonnet et al. (2012), demonstrating that prior knowledge of both conceptual and grammatical gender information is strategically made to use by bilinguals in order to make categorical judgments about gender. However, the results of Experiment 1 do not allow us to determine the extent and strength to which grammatical and conceptual gender information permeate non-verbal cognitive processes. In other words, it remains unclear as to whether their impact emerges even in cases when the information is unnecessary for the immediate task. To disentangle these effects, Experiment 2 presented participants with a perceptual judgment task on a target genderless face, in which the object primes were irrelevant for making their decisions.

3. Experiment 2

3.1. Method

3.1.1. Participants

Twenty-seven monolingual native English speakers (18 women; M_{age} = 22.56, age range = 18 – 50 years) and 28 French-English bilinguals (14 women; M_{age} = 24.21, age range = 17 – 50 years) from Lancaster University (U.K.) were paid to participate in the study. All participants gave their informed written consent. Information about the L2 background of the bilingual participants is indicated in Table 1. Comparable to Experiment 1, the French-English bilinguals scored on average 76.81% (*SD* : 14.79) on the QPT which is equivalent to a C1 or advanced level on the CEFR. Self-evaluation of the bilinguals' L2 English in speaking, reading and understanding of spoken language was, again, evaluated as adequate or above. All participants had normal or corrected to normal vision. None of the participants had taken part in Experiment 1.

3.1.2. Materials and Design

The materials were identical to that of Experiment 1 with the exception that all target facial images presented at the end of the triads were replaced with one of the eight genderless facial images normed and implemented in Slepian, Weisbuch, Rule, and Ambady (2011). Unlike Experiment 1 where the experimental conditions were dependent on the combination of the object gender and target face sex, the experimental conditions in Experiment 2 were based solely on the conceptual gender association and grammatical gender class of each object prime (see Table 4 for a detailed description of the experimental conditions and Appendix A: Table A.1 for all items per condition). These yielded conditions where the object prime was: (1) conceptually gender-neutral and was grammatically (1a) feminine or (1b) masculine in French, (2) conceptually female, and was grammatically (2a) feminine or (2b) masculine in French, or (3) conceptually male and was grammatically (3a) feminine or (3b) masculine in French. The genderless facial images were each displayed in the middle of the screen alongside a female and male trait word on each of its sides.

Table 4.

Experimental conditions and their corresponding gender combination of items in Experiment 2

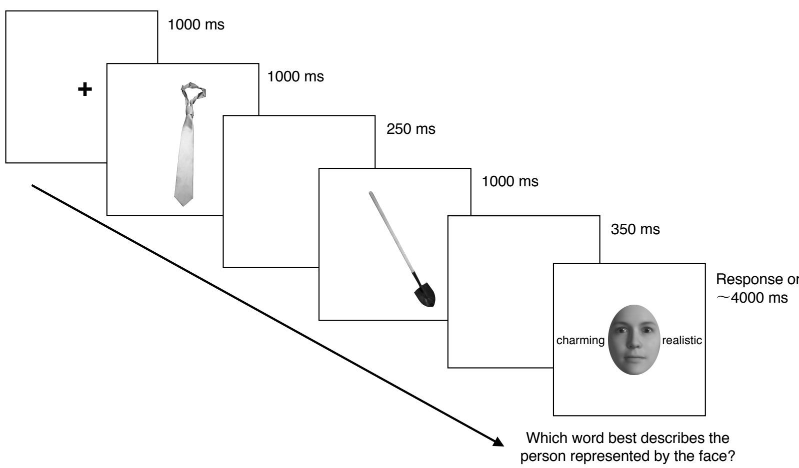
Experimental Conditions		Target face sex
Conceptual gender of the object prime	Grammatical gender of the object prime	
Neutral	(1a) *Feminine*	Genderless
	(1b) *Masculine*	
Female	(2a) *Feminine*	Genderless
	(2b) *Masculine*	
Male	(3a) *Feminine*	Genderless
	(3b) *Masculine*	

Six typically-female and six typically-male gendered trait words were selected from a gender-normed list by Archer and Lloyd (2002). Six pairs of gendered trait words were formed, each which consisted of a female and male trait word of similar valence (neutral, positive and negative), frequency (as measured by SUBTLEX-UK: Heuven, Mandera, Keuleers, & Brysbaert, 2014), and word length (See Appendix A : Table A.2 for a full list of the details of the gendered trait word pairs). Word lengths differed by a maximum of one letter within the pairs (e.g., female trait: *charming*; male trait: *realistic*). Each experimental item was presented four times to counterbalance the presentation order of the first two object images and the physical position in which the trait words appeared next to the face (left or right of the face). Filler items consisted of the same items presented in Experiment 1.

3.1.3. Procedure

Conditions of stimuli presentation were identical to that of Experiment 1. Upon the presentation of the genderless face, participants were instructed to decide as accurately as possible which of the two trait words best described the target face. The instructions did not pressure the participants to make quick responses as the trait word pairs differed in word length and frequency between the pairs, although the succeeding trial was presented after 4000 ms elapsed. A detailed timeline of the experimental procedure and example stimuli is shown in Figure 4.

Figure 4. Experimental timeline for critical trials in Experiment 2. A fixation point was presented for



1000 ms, followed by each object presented consecutively for 1000 ms with an inter-stimulus interval of 250 ms between the objects. After a 350 ms interval following the second object, a genderless face image was presented. Participants decided which of the two gendered trait words presented alongside the face best described the face with a key press on the keyboard corresponding to the physical orientation in which the trait word appeared on the computer screen (i.e., left or right of the face).

Absence to respond consisted only of .01 % of the native English speakers' responses and .03 % of the French-English bilinguals' responses to experimental items. For this task, no explicit link needed to be made between the prime and target face, but the object primes needed to be perceived in order to respond to the filler trials. Responses to the experimental items were collected with a key press on the keyboard corresponding to the physical orientation in which the trait word appeared on the computer screen (i.e., left or right of the face). Instructions for the filler items were identical to that of Experiment 1. Yes / no responses to filler items were also allocated the same keys, with the positive response being designated to the participants' dominant hand as was the case in Experiment 1. Four practice items preceded the main experimental task, and the bilinguals finished the session with the completion of the LEAP-Q (Marian et al., 2007) and QPT (Oxford University Press, 2001).

3.2. Results

As participant responses were binary (i.e., female or male trait word), mixed-logit models were fitted to predict participants' likelihood of designating a female or male trait word to a genderless face (see Figure 5 for mean proportions of trait word allocations). Female trait words were coded as hit responses (i.e., 1 = female trait word, 0 = male trait word) as past studies have indicated that the male gender may commonly reflect a default gender during language processing (e.g., Irmen, 2007; Silveira, 1980). The fixed effects consisted of the *conceptual gender* associated with the first two object primes (3: neutral vs. female vs. male), their *grammatical gender* category of the objects in French (2: masculine vs. feminine), and their interaction. We followed the same modeling conditions as Experiment 1, including a maximal random effects structure for both items and participants.

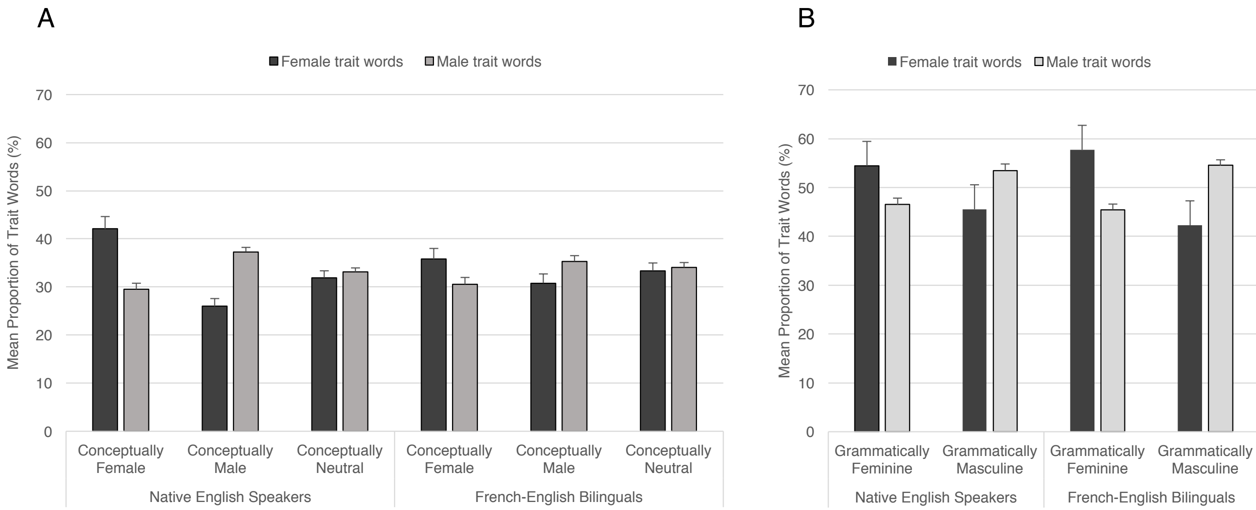


Figure 5. Mean proportions of trait words allocated for each (A) conceptual gender and (B) grammatical gender condition per language group. Error bars indicate standard error of the mean.

Similar to Experiment 1, we assumed that the acceptance or rejection of activated conceptual gender would not differ for gender-specific triads (i.e., male and female conceptual gender). Thus, conceptual gender was coded with orthogonal contrasts, with the first comparing the effect of neutral gender to that with a specific gender (average of male and female conceptual gender), and the second contrast comparing the conceptual gender effect of male to that of the female gender. For grammatical gender class, sum contrasts were applied (grammatically feminine: 0.5, grammatically masculine: -0.5). The fixed effects structure of the final model is summarized in Table 5. In the text, we report back-transformed odds ratios of the parameter estimates for better interpretability.

	Native English speakers				*French-English bilinguals*			
	Estimate	*SE*	*z*	*p*	*Estimate*	*SE*	*z*	*p*
(Intercept)	- 0.47	0.12	-3.76	< .001	- 0.48	0.13	- 3.44	< .001
CG (M and F vs. N)	0.03	0.14	0.24	*ns*.	0.06	0.25	0.25	*ns*.
CG (F vs. M)	0.71	0.16	4.5	< .001	0.38	0.33	1.16	*ns*.
GG (FG vs. MG)	0.27	0.17	1.56	*ns*.	0.54	0.21	2.51	< .01
CG (M and F vs. N) : GG (FG vs. MG)	- 0.48	0.27	- 1.77	*ns*.	- 0.22	0.46	- 0.48	*ns*.
CG (F vs. M) : GG (FG vs. MG)	- 0.29	0.32	- 0.91	*ns*.	- 0.04	0.53	- 0.08	*ns*.

Note: CG= Conceptual Gender; GG = Grammatical Gender; N = Conceptually Neutral; F = Conceptually Female; M = Conceptually Male; MG = Grammatically Masculine; FG = Grammatically Feminine

Table 5. *Fixed effects estimates of mixed logit model of gendered trait word* allocation *for each language group (Experiment 2)*

3.2.1. Native English speakers

The final model for the native English speakers included by-participant and by-item intercepts, as well as by-participant and by-item random slopes for grammatical gender. The first contrast for conceptual gender revealed that there was no significant difference between the assignment of trait words following conceptually gender-neutral and gender-specific objects (*ns.*). However, the second contrast showed that participants' odds of allocating female trait words increased 2.03 times when the faces followed conceptually female than conceptually male gender-associated objects ($p < .01$). As was the case for Experiment 1, native English speakers reliably identified the conceptual gender association of the objects, which was used to allocate a gender-associated trait word to a genderless face. These effects emerged even though the task did not require any explicit associations to be made between the object prime and target face. No other effects were significant (*ns.*).

3.2.2. French-English bilinguals

The final model for the French-English bilinguals included by-participant and by-item intercepts, as well as by-participant slopes for conceptual gender. Contrary to the results of Experiment 1, the model revealed that the conceptual gender association of the objects did not predict the allocation of gender-associated trait words to the target face (both contrasts: *ns.*). Crucially for the study, the effect of grammatical gender was significant, indicating that when the objects were grammatically feminine as opposed to masculine, participants' odds of allocating female traits increased 1.72 times ($p < .01$). Importantly, this effect did not interact with the conceptual gender of the objects (*ns.*).

3.3. Discussion

Whereas participants in Experiment 1 were required to consciously identify task-relevant information to make their judgments, in Experiment 2, this was not the case. There were neither references as to how the objects related to the target face nor requirements to link the object primes to the target face in any way. Participants were instructed to simply focus on selecting the trait word that best defined the genderless face, and the presentation of the objects was only intended for the filler items. Even under the absence of such contextual cues, prior knowledge about conceptual gender associations reliably predicted the native English speakers' categorical judgments. In contrast, the French-English bilinguals did not activate conceptual gender. Instead, grammatical gender determined participants' perceptual judgments of sex-related traits of genderless faces. Note that the elicitation of judgments was not based on explicit male or female choices but was only nuanced through an implicit task of allocating a gender-associated trait word. These results suggest that simply being exposed to an object appears to activate gender features that may facilitate the resolution of the task. Provided that the bilinguals only activated grammatical gender information, it appears to offer a more robust source that overrides conceptual gender information whilst implicitly attending to sex-related information.

4. General Discussion and Conclusion

The present study aimed to take first steps in providing evidence for the view that discrete grammatical properties offer structural feedback which may lead to perceptual biases during categorization. Across two experiments, we found that native English speakers consistently and exclusively relied on conceptual gender, even if this information was contextually insignificant (Experiment 2). As for the French-English bilinguals, the effect of conceptual gender emerged only in conditions where this information was related to the task (Experiment 1). In fact, they consistently recruited the grammatical gender information of the objects which successively biased their judgments regarding sex-related information. Specifically, grammatical gender continued to influence responses even when such information was task-irrelevant (Experiment 2). Simply being exposed to objects appeared to have activated their grammatical gender category, suggesting that grammatical gender provides a much more immediate and robust effect on non-verbal categorization.

The variability in the effects arising from the two experiments speaks to the differences in how grammatical and conceptual information may be retrieved. When participants deduced that conceptual gender could be extracted to strategically resolve the task (Experiment 1), they made necessary gender associations about the objects based on their prior knowledge. In contrast, when no contextual cues were available to guide their judgments (Experiment 2), the ambiguity was substituted with resources that were readily accessible (e.g., Montero-Melis, Jaeger, & Bylund, 2016). For the native English speakers, this information was conceptual gender which was primed immediately before the elicitation of the judgment. The French-English bilinguals, on the other hand, did not draw on conceptual gender but based their judgments exclusively on grammatical gender. A probable explanation for these results is that retrieving conceptual gender entails a conscious focus on gender relations where perceivers generate inferences taking into account contextual constraints (e.g., task relevance) and prior knowledge. In this regard, conceptual gender may or may not be activated. However, activating grammatical information is requisite during the representation of an object because gender is an obligatory grammatical category. Not only is grammatical gender independent of contextual factors, but it is also accessed on an unconscious and uncontrollable level (Boutonnet et al., 2012). Crucially, the observed grammatical gender effect in both experiments was elicited from bilinguals who were tested solely in an English-speaking environment. Although it would seem plausible to observe a similar effect had we tested monolingual French speakers in a French-speaking context (in which case the effect would most likely have been stronger), the conditions in which the grammatical effects surfaced in our study speaks to the potency of its influence.

These results suggest that knowledge of grammatical gender may provide a strong structural basis which is recruited on-line, motivating perceptual biases to emerge. Given that the participants were given minimal information to base their judgments in Experiment 2, the bilinguals' reliance on available knowledge about grammatical categories would have been inevitable. Our findings concur with the general notion to which the label-feedback hypothesis puts forward: the unconscious activation of linguistic information occurs during on-line categorical perception, and in so doing, flexes perception by means of emphasizing associated features. Our study further complements this view by theorizing the precise effects of grammatical features on cognition. Specifically, we argue that the learned associations that facilitate such an effect are not constrained to whole labels, rich in information about a category's feature properties, but extend even onto grammatical categories. Such a view resonates with a point made by Lucy, (1992) on the effects of grammar on thought, but seldom empirically followed up, as the majority of linguistic relativity studies have focused on the lexical domain (e.g., color).

The mechanism guiding this robust effect can be accounted for by theories of associative learning (e.g., L. B. Smith & Samuelson, 2006) which predict that concept formation of an object requires perceivers to learn that it frequently occurs within a specific context or is accompanied by specific linguistic or perceptual properties (Bylund & Athanasopoulos, 2014). Because identifying these statistical tendencies drives perceivers' attention to these diagnostic properties, cognitive categorization may consequently bear an alignment to these features. In terms of our study, an object's representation is inherently contingent on the access to its grammatical gender information. Although they may not necessarily carry semantically-relevant information for the immediate context, this learned association may result in playing a central role in perceptual processes. It is not surprising if years of experience in attending to these categorical encodings would have predisposed our bilingual speakers with a structural bias of their judgments about sex-related information. Moreover, the fact that speakers of grammatical gender languages are cognitively more committed to these grammatical categories than to specific lexical entries lends the possibility that grammar may have more durable and stronger consequences on cognition. This would imply that morphological cues may suffice in modulating perceptual representations and that whole categorical labels are not the only form of linguistic knowledge affecting our interpretation of reality. Our findings thus support the linguistic relativity hypothesis (Whorf, 1956) showing the way we interpret reality and make evaluative judgments of perceptual stimuli very much depends on the grammatical categories of the languages we use.

Although one may argue that the use of linguistic materials (i.e., trait words) in Experiment 2 may have induced the French-English bilinguals to rely on language to resolve the task (i.e., the effect of 'language on language': Casasanto, 2008; Lucy, 1997), we note that all experimental manipulations investigating linguistic relativity effects rely on language to some extent, be it for giving instructions, or for the presentation of experimental stimuli to execute the task (e.g., Boroditsky et al., 2003; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013). More importantly, we argue that our experimental task was markedly implicit, which minimized the possible overt link between the language used and the specific grammatical feature investigated. Judgments on the target faces were completely disassociated from the object primes, and the designation of sex was conducted simply by allocating a gender-trait word to a genderless face. In fact, the trait words were not only presented in English, but their translation in French cannot account for the grammatical inflection of a specific gender. The trait words were thus ambiguous with regard to any grammatical links, and yet, the data strongly substantiate our reasoning indicating that the two language groups relied on different kinds of information sources to complete the task. Crucially, during the debriefing, not a single participant irrespective of language background reported that they had been aware that the task was designed to elicit their judgments regarding the sexual properties of the target face.

We note, however, that questions remain open as to the extent these grammatical gender effects may account for cases where bilinguals have acquired two separate grammatical gender languages. Past studies implementing gender assignment tasks have found that even in the presence of two grammatical gender categories, bilinguals nonetheless show sensitivity to both grammatical systems (e.g., Spanish-German bilinguals: Boroditsky et al., 2003). Together with these findings, our study provides reason to assume that similar effects may surface. However, future studies may explore the possible interactions between bilinguals speaking different grammatical gender languages.

In sum, our study lends further support for the view that grammatical properties are automatically retrieved during the perceptual representation of objects. The study offers additional findings to ascertain that grammatical information provides a strong structural basis that may subsequently influence categorical judgments. As indicated by accounts of predictive processing, its activation provides enhanced access to general gender-related information which may facilitate or bias the representation of sex-associated information. While non-verbal prior knowledge (i.e., conceptual gender) may also guide categorization, the impact of grammatical information appears to be inevitable and permeates into cognitive processes unconsciously. The present study provides a first look at the extent to which different sources of prior knowledge modulate categorical decisions about the sex of human faces. Future studies may implement more implicit neurophysiological measures such as event-related brain potentials that would shed light on the extent of the cognitive penetrability of conceptual and grammatical gender in early visual processing (e.g., Boutonnet & Lupyan, 2015; Flecken et al., 2015; Thierry et al., 2009). For now, our findings indicate that grammar enables an effective means to access prior knowledge when making perceptual judgments which may be independent of task requirements.

Acknowledgments

This work was supported by the Swiss National Science Foundation [grant number P2FRP1_161712].

References

Archer, J., & Lloyd, B. B. (2002). *Sex and gender* (2nd ed.). Cambridge, UK: Cambridge University Press.

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*, 89–96. https://doi.org/10.1017/S1366728905002397

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*, 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., Bylund, E., & Casasanto, D. (2016). Introduction to the special issue: New and interdisciplinary approaches to linguistic relativity. *Language Learning*, *66*, 482–486. https://doi.org/10.1111/lang.12196

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., … Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, *26*, 518–526. https://doi.org/10.1177/0956797614567509

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*, 1–28.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79. https://doi.org/10.1016/j.brainres.2012.08.044

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience*, *35*, 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLOS ONE*, *5*, e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Kehayia, E., Dion-Lessard, G., Chauret, M., Montreuil, T., Dionne-Dostie, E., & Lepage, M. (2012). The Bank of Standardized Stimuli (BOSS): Comparison between French and English norms. *Behavior Research Methods*, *44*, 961–970. https://doi.org/10.3758/s13428-011-0184-7

Bylund, E., & Athanasopoulos, P. (2014). Linguistic relativity in SLA: Toward a new research program. *Language Learning*, *64*, 952–985. https://doi.org/10.1111/lang.12080

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

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*, 579–593. https://doi.org/10.1016/j.cognition.2007.03.004

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

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*, 449–460. https://doi.org/10.1037/a0021965

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*, 203–204. https://doi.org/10.1038/18335

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974. https://doi.org/10.1126/science.284.5416.970

Fausey, C. M., & Boroditsky, L. (2010). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*, 150–157. https://doi.org/10.3758/s13423-010-0021-5

Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology*, *32*, 18–31. https://doi.org/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. https://doi.org/10.1016/j.cognition.2015.04.006

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*, 819–824. https://doi.org/10.1016/j.cognition.2008.04.007

Heuven, W. J. B. van, Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2014). All giraffes have female-specific properties: Influence of grammatical gender on deductive reasoning about sex-specific properties in German speakers. *Cognitive Science*, *38*, 514–536. https://doi.org/10.1111/cogs.12074

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*, 431–456. https://doi.org/10.1007/s10936-007-9053-z

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

Kovács, G., Gulyás, B., Savic, I., Perrett, D. I., Cornwell, R. E., Little, A. C., … Vidnyánszky, Z. (2004). Smelling human sex hormone-like compounds affects face gender judgment of men. *Neuroreport*, *15*, 1275–1277. https://doi.org/10.1097/01.wnr.0000130234.51411.0e.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models (Version 2.0-32). Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge University Press.

Lucy, J. A. (1997). Linguistic relativity. *Annual Review of Anthropology*, *26*, 291–312. https://doi.org/10.1146/annurev.anthro.26.1.291

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Clark, A. (2015). Words and the world predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*, 279–284. https://doi.org/10.1177/0963721415570732

Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS ONE*, *5*, 1–9. https://doi.org/10.1371/journal.pone.0011452

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*, 2236–2253. https://doi.org/10.3758/BF03196698

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology. General*, *141*, 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*, 14196–14201. https://doi.org/10.1073/pnas.1303312110

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114. https://doi.org/10.1016/j.tics.2004.01.003

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Montero-Melis, G., Jaeger, T. F., & Bylund, E. (2016). Thinking is modulated by recent linguistic experience: Second language priming affects perceived event similarity. *Language Learning*, *66*, 636–665. https://doi.org/10.1111/lang.12172

Oxford: Oxford University Press. (2001). Oxford Quick Placement Test. Oxford University Press.

Pinker, S. (1994). *The language instinct*. London: Penguin.

R Core Team. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369–398. https://doi.org/10.1037/0096-3445.129.3.369

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, M.A.: MIT Press.

Saalbach, H., Imai, M., & Schalk, L. (2012). Grammatical gender and inferences about biological properties in German-speaking children. *Cognitive Science*, *36*, 1251–1267. https://doi.org/10.1111/j.1551-6709.2012.01251.x

Sato, S., Gygax, P. M., & Gabriel, U. (2016). Gauging the impact of gender grammaticization in different languages: Application of a linguistic-visual paradigm. *Frontiers in Psychology*, 140. https://doi.org/10.3389/fpsyg.2016.00140

Sera, M. D., Berge, C. A. H., & del Castillo-Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*, 261–292. https://doi.org/10.1016/0885-2014(94)90007-8

Silveira, J. (1980). Generic masculine words and thinking. In C. Kramarae (Ed.), *The voices and words of women and men* (pp. 165–178). Oxford, U.K.: Pergamon.

Slepian, M. L., Weisbuch, M., Rule, N. O., & Ambady, N. (2011). Tough and tender: Embodied categorization of gender. *Psychological Science*, *22*, 26–28. https://doi.org/10.1177/0956797610390388

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology: CB*, *17*, 1680–1685. https://doi.org/10.1016/j.cub.2007.08.043

Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology*, *42*, 1339–1343. https://doi.org/10.1037/0012-1649.42.6.1339

Thierry, G. (2016). Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning*, *66*, 690–713. https://doi.org/10.1111/lang.12186

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*, 4567–4570. https://doi.org/10.1073/pnas.0811155106

Utz, S., & Carbon, C.-C. (2015). Afterimages are biased by top-down information. *Perception*, *44*, 1263–1274. https://doi.org/10.1177/0301006615596900

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

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*, 7780–7785. https://doi.org/10.1073/pnas.0701644104