The Influence of Three-Gendered Grammatical Systems on Simultaneous Bilingual Cognition: The case of Ukrainian-Russian Bilinguals

Abstract

This paper examines the linguistic relativity principle (Whorf, 1956) by investigating the impact of grammatical gender on cognition in simultaneous bilinguals of three-gendered Ukrainian and Russian. It examines whether speakers of three-gendered languages show grammatical gender effects on categorisation, empirically addressing claims that such effects are insignificant due to the presence of the neuter gender (Sera et al., 2002). We conducted two experiments using a similarity-judgment paradigm while manipulating the presence of neuter gender stimuli (Phillips & Boroditsky, 2003). Experiment 1, including neuter gender, revealed no significant effects, compatible with earlier studies on three-gendered languages. Conversely, Experiment 2, excluding neuter gender stimuli, showed significant language effects. Bilingual participants rated pairs as more similar when grammatical genders in both languages were congruent with the biological sex of a character. Significant effects were also found for pairs with mismatching grammatical genders in Ukrainian and Russian. Participants with higher proficiency in Ukrainian rated pairs as more similar when the grammatical gender of a noun in Ukrainian was congruent with the character's biological sex, and incongruent in Russian. Our findings thus provide the first empirical demonstration that the exclusion of neuter gender online induces grammatical gender effects in speakers of three-gendered languages.

Keywords: linguistic relativity, grammatical gender, simultaneous bilingualism, language proficiency

Authors:

1. **Oleksandra Osypenko**
   Department of Linguistics and English Language,
   Lancaster University,
   Lancaster, United Kingdom
   Email: o.osypenko@lancaster.ac.uk
2. **Silke Brandt**
   Department of Linguistics and English Language,
   Lancaster University,
   Lancaster, United Kingdom
   Email: s.brandt@lancaster.ac.uk
3. **Panos Athanasopoulos**
   Centre for Languages and Literature,
   Lund University,
   Lund, Sweden
   Email: panos.athanasopoulos@englund.lu.se

Corresponding Author:

Oleksandra Osypenko
Department of Linguistics and English Language,
County South, Lancaster University,
Lancaster, LA1 4YL, United Kingdom
Email: o.osypenko@lancaster.ac.uk

1. Introduction

The majority of studies investigating linguistic relativity effects typically concentrate on the question "Does language influence our thoughts?" (Athanasopoulos & Casaponsa, 2020). While this question has been asked in a number of disciplines, such as philosophy, linguistics, anthropology and psychology, modern versions of the question can be traced to Whorf (1956) and more recent trans-disciplinary scholarly activity (Lucy, 1997; Gentner & Goldin-Meadow, 2003; Athanasopoulos, Bylund & Casasanto, 2016), which has placed the question at the forefront of cognitive science. Various domains have been used as a testbed for the hypothesis, such as spaciotemporal metaphors (Athanasopoulos & Bylund, 2023), colour (Athanasopoulos, 2009; Winawer et al., 2007), and grammatical gender (Boroditsky & Schmidt, 2000; Boutonnet et al., 2012; Sato & Athanasopoulos, 2018).

The latest surge of attention led to more detailed explanations of the effects languages may have on cognitive processes, by including various experimental conditions, such as verbal interference, differentiating stimuli based on their perceptual characteristics, or manipulating the complexity of experimental design (Athanasopoulos & Casaponsa, 2020). Therefore, posing the aforementioned question as one that requires a binary answer seems out-dated. Instead, the focus is moving away from providing evidence to a "yes-no" question towards investigating what circumstances lead to emerging language effects on cognitive processes (e.g., memory or categorisation), as well as how and why language-specific features form the groundwork for individual perceptual judgement, including multilingual speakers (Bassetti & Filipović, 2022; Casasanto, 2016). An illustrative example of the latter in our study pertains to the emergence of grammatical gender effects in speakers of three-gendered languages. Previous research on linguistic relativity (Sera et al., 2002; Vigliocco et al., 2005) has reported the absence of such effects, while more recent studies yield mixed results (Pavlidou & Alvanoudi, 2019). The primary factor contributing to these mixed or non-

emergent outcomes has been hypothesised to be the presence of the neuter gender in these languages, which is thought to diminish the prominence of gender effects. Consequently, our research seeks to determine whether grammatical gender effects on cognitive processes, such as categorisation, are confined to two-gendered languages or can also be observed in speakers of three-gendered languages, and under what specific conditions these effects manifest.

We also focus on bilingual speakers who have two partially conflicting grammatical systems (where some nouns have matching and others mismatching grammatical gender in Ukrainian and Russian). Specifically, the impact two grammatical gender systems have on perception and categorisation, even when participants are not actively engaging with either language, as the testing was conducted entirely in English, which unlike Russian and Ukrainian does not have a grammatical gender system.

Generally, research on language and cognition in bilinguals continues to be an important endeavour of the linguistic relativity theory complex, as Whorf (1956) himself pointed out that if language affects our thoughts, then learning other languages can free people from the shackles of their own language. Employing Ukrainian-Russian simultaneous bilinguals is of interest because the representation of two grammatical gender systems within an individual's mind and their effects on bilinguals' cognitive processes, such as memory or categorisation, have received little attention (e.g., the study by Bassetti, 2007). It remains unclear whether language effects would emerge only when grammatical gender matches in both languages or if they would also occur when grammatical gender mismatches, depending on the more proficient language. Additionally, there is uncertainty whether any effects would appear at all, given that both languages include a neuter gender in their grammatical system.

Here, we attempt to investigate the effects that two partially contrasting three-gendered grammatical systems (e.g., Ukrainian as L1 and Russian as 2L1) have on categorisation, as well as introduce simultaneous bilinguals with two distinct grammatical

gender systems into linguistic relativity research. In addition, at a theoretical level, we aim to explore whether the presence of neuter grammatical gender mitigates language effects, as suggested previously (Sera et al., 2002; Vigliocco et al., 2005). To do so we employed a similarity judgement paradigm while manipulating stimuli with (Experiment 1) and without neuter gender (Experiment 2). Such manipulation would also allow us to investigate further into the nature of the gender effects, particularly whether (if found) the effects of grammatical gender arise online (in the moment of testing) or offline (entrenched in previous language experience) (Lupyan, 2020). If the effects arose online (Lupyan, 2012; Sato & Athanasopoulos, 2018), we anticipated observing more pronounced effects in Experiment 2, whereas if the effects were offline, comparable effects were expected across both Experiment 1 and Experiment 2.

1.1. Grammatical gender in language and mind

The empirical evidence of linguistic relativity effects can be found across various domains, such as colour categorisation/discrimination (Athanasopoulos, 2009; Roberson et al., 2005; Winawer et al., 2007), time and space (Athanasopoulos & Bylund, 2023; Boroditsky, 2001; Casasanto et al., 2004), motion (Athanasopoulos & Bylund, 2013), grammatical number and object classification (Athanasopoulos, 2006; Lucy, 1992), tactile perception (Miller et al., 2018) and even olfaction (Cao et al., 2024; Speed & Majid, 2019; Vanek et al., 2021). This evidence supports the idea that the structure of language can shape non-linguistic cognition, offering a compelling testbed for investigating how grammatical features, such as gender, influence thought.

Grammatical gender has been used as a subject of analysis by linguistic relativity researchers because of two primary reasons. Firstly, when grammatical gender is absent, no other lexicalisation pattern can replace it (Boutonnet et al., 2012). Secondly, the assignment

of grammatical gender to inanimate nouns, and certain animals in the case of Ukrainian and Russian, is usually unpredictable and semantically illogical (Elpers et al., 2022). For instance, “parrot” in Ukrainian takes the feminine grammatical gender, while in Russian it is masculine. Besides, even though grammatical gender is superfluous for interaction in the case of many languages (e.g., English), for speakers of various languages, such as Russian and Ukrainian, it cannot be ignored. In such languages the gender of objects is mandatorily marked in a range of morphosyntactic constructions, such as demonstratives, pronouns, singular adjectives, and verbs in the past tense (Mitrofanova et al., 2018). Such morphosyntactic consequences of grammatical gender make it an ideal candidate for examining whether grammatical categories influence cognitive processes beyond lexical features (Sato & Athanasopoulos, 2018).

Despite extensive research, a notable gap exists in understanding the cognitive effects of grammatical gender across different grammatical systems, particularly three-gendered languages. Most studies have focused on German (Bassetti, 2007; Pavlidou & Alvanoudi, 2019; Sera et al., 2002; Vigliocco et al., 2005), which may yield less significant results due to inconsistencies in gender assignment (e.g., “das Mädchen” [a girl] being neuter) and the use of articles that do not always differentiate between genders (e.g., the dative case where both masculine and neuter use “dem”). In contrast, Ukrainian and Russian, both three-gendered, indicate gender primarily through noun endings, providing a more consistent gender-marking system. By extending research to these underrepresented languages, this study aims to offer new insights into how three-gendered grammatical systems influence cognitive processes.

A wide range of behavioural tasks have been developed to study the impact of grammatical gender on cognitive representation of concepts, with the most common one being the voice attribution task (i.e., asking participants to assign either a male or female voice to objects; see Samuel et al., 2019). Other methods include a sex assignment task

(Belacchi & Cubelli, 2012), an object-name memory task (Boroditsky & Schmidt, 2000), and a similarity judgement task (Phillips & Boroditsky, 2003). The current study employs the similarity judgment task, where participants rate the similarity between pairs of depicted objects and characters with a clear biological sex using a Likert scale. The choice of this paradigm is rooted in its unique strengths, such as it requires using unlabelled stimuli that minimise active language processing that is a key element in testing whether language shapes non-linguistic representations (Casasanto, 2016). This methodology was first implemented in linguistic relativity research in in the seminal work of Phillips & Boroditsky (2003), who argued that Spanish-English and German-English sequential bilinguals perceived object-personified character pairs as more similar when the biological sex of the character and the grammatical gender of the object in their L1 were congruent, even when tested in English. This suggests that grammatical gender influences object categorisation even when grammatical gender is not explicitly used. Overall, the research has shown that when making gender-related judgments, individuals often take into account the object's grammatical gender (Flaherty, 2001; Konishi, 1993). Despite more recent studies that produced contrasting results and highlighted the issue of a replication crisis, including a failed replication by Elpers et al. (2022) and mixed findings by Sedlmeier et al. (2016), the study by Phillips and Boroditsky (2003) has nonetheless made a significant impact on the field.

One possible explanation for the mixed findings might be linked to the type of grammatical gender system present in a language, particularly the distinction between two-gendered and three-gendered systems. For instance, Sera et al. (2002) found that, unlike Spanish and French monolingual children, German children did not use grammatical gender to assign voices to objects during categorisation tasks, instead aligning their responded more closely to Spanish gender. The study suggests that two-gendered languages have a stronger association between grammatical and natural gender, leading to overgeneralisation of

masculine and feminine traits to inanimate objects. In contrast, speakers of languages with a three-gender system, such as German, appear to rely less on gender and more on other conceptual distinctions when categorising objects. Similarly, Vigliocco et al. (2005) found significant gender effects in Italian but not in German during a similarity judgment task, arguing that the weaker link between grammatical gender and semantic properties in three-gender systems results in reduced gender effects on perception. Inconsistencies in gender assignment and a lack of clear correspondence with the sex of referents likely contribute to this difference. The authors suggest that the mapping between grammatical gender and semantic properties is weaker in three-gender systems like German compared to two-gender systems like Italian. They argue that three-gendered languages do not exhibit the same grammatical gender effects because the correspondence between gender and the sex of referents is less transparent. To address these criticisms and further examine the role of grammatical gender in three-gender systems, Pavlidou and Alvanoudi (2019) conducted a sex-attribution task (adapted from Sera et al., 2002) with speakers of German and Greek (both three-gendered languages). Participants were asked to assign names to depicted nouns for a preschool play, with nouns having masculine, feminine, or neuter gender. Their analysis revealed significant effects of grammatical gender on sex-attribution in both languages, challenging earlier claims by Sera, et al. (2002) and Vigliocco, et al. (2005).

Similarly, Bassetti (2007) – the only study to our knowledge that examined simultaneous bilinguals when looking at grammatical gender effects in linguistic relativity research - investigated how grammatical gender influences categorisation and representations of concepts in Italian-German simultaneous bilingual and Italian monolingual children using a voice attribution task. This is particularly relevant to the current study because objects were also chosen with opposite genders in Italian and German. Results showed that grammatical gender effects were only present in Italian monolinguals, echoing Sera et al. (2002),

suggesting that Italian gender assignment may be more intuitive or 'natural' compared to German. The study also noted that bilinguals, who navigate two languages with mismatched grammatical gender systems, develop unique cognitive frameworks, integrating elements from both languages. Consequently, bilinguals may think differently from monolinguals, not because of bilingualism itself, but due to the specific characteristics of the grammatical systems embedded in the languages they speak, such as mismatching grammatical genders in Italian and German. This observation is particularly relevant to our study, as we also examine partially mismatching grammatical gender systems, albeit within two three-gendered languages.

1.2. Online vs offline nature of the grammatical gender effects

A central question in this line of research is whether grammatical gender effects operate online (as real-time, context-sensitive influences) or offline (as enduring impacts of long-term linguistic experience). According to Lupyan et al. (2020), online effects occur when language actively modulates perception and decision-making in the moment, often shaped by top-down feedback from linguistic labels and grammatical structures. Offline effects, in contrast, reflect long-term, habitual patterns ingrained by extensive language use that influence perception even outside linguistic contexts.

This study draws on two complementary theoretical frameworks to address this distinction. The label-feedback hypothesis (Lupyan, 2012) proposes that even when no explicit labels are presented, internal labelling processes may still influence perception and categorisation in real time. This reflects a top-down influence, where prior language knowledge actively shapes what features are noticed or emphasised during perception. Extending this idea, the structural-feedback hypothesis (Sato & Athanasopoulos, 2018) posits

that the influence of grammatical gender extends beyond specific labels, stemming from the broader habitual patterns ingrained by the grammatical system itself. According to this hypothesis, grammatical gender activates unconsciously during the online categorical perception and by doing so, it modulates perception by emphasising the features associated with it.

The current study aims to directly engage with the online vs offline debate by designing two similar experiments with the main difference being that Experiment 1 includes objects of all three grammatical genders (masculine, feminine, and neuter), while Experiment 2 excludes neuter gender. This allowed us to test whether the presence of neuter stimuli dilutes the salience of masculine-feminine distinctions, potentially weakening online grammatical gender effects. If the effects are online, we expect stronger effects in Experiment 2, as removing neuter gender heightens the binary masculine-feminine distinction. Conversely, if the effects are offline, results should remain consistent across both experiments, reflecting the enduring impact of long-term linguistic patterns rather than immediate task context.

To sum up, given the mixed results demonstrated in studies involving speakers of three-gendered languages, it is important to note that no previous research has directly compared the strength of grammatical gender effects using the same task with and without the inclusion of neuter gender. The present study uniquely investigates the cognitive effects of bilingualism in two conflicting three-gendered languages, a topic that has not been previously explored. Besides, we extend research beyond typically used German to other three-gendered languages (Ukrainian and Russian). This approach provides a more comprehensive understanding of how grammatical gender influences cognition across diverse linguistic contexts.

1.3. Case of Ukrainian simultaneous bilingualism and typological differences in Ukrainian and Russian languages

Simultaneous bilingualism in Ukraine presents unique challenges and insights into the cognitive processing of language, particularly when the languages involved have distinct grammatical systems. This study focuses on Ukrainian-Russian bilingualism, specifically the typological differences between the languages, especially regarding grammatical gender.

Ukraine has a deep-rooted history of multilingualism (Poftak & Shykula, 2022), and the status of the Russian language has long been a subject of debate (Eberhard et al., 2019). According to the 2001 census, out of Ukraine's then-population of 48.5 million, 78% identified as Ukrainians and 17% identified as Russians when asked to choose one ethnic affiliation. However, linguistic preferences differed, with 68% selecting Ukrainian as their native language and 30% opting for Russian (Bilaniuk & Melnyk, 2008). Despite the historical stigmatisation of bilingualism even prior to the war (Pavlenko, 2012), it is clear that societal bilingualism is inherent in Ukraine (Csernicskó & Máté, 2017; Shumlianskyi, 2010).

The onset of the war in February 2022 dramatically altered these linguistic landscapes. There has been a sharp increase in the proportion of respondents who, according to self-reported questionnaires, speak predominantly Ukrainian in everyday life and a corresponding decrease of Russian speakers. The most recent poll from December 2022 indicates that 41% of respondents claimed to communicate only in Ukrainian, another 17% reported using Ukrainian "in most situations", while only 6% speak only in Russian, and 9% predominantly in Russian, another 24% said they use both languages "equally" (Kulyk, 2023). Compared to 2017, the proportion of exclusive and predominant Ukrainian speakers increased by 8%, and the proportion of Russian speakers decreased by 11% (Kulyk, 2023). Given the fluid language attitudes and shifting language use among bilingual individuals in

Ukraine, it is worth examining which languages have the most significant impact on cognitive processes of such speakers. It has been proposed in linguistic relativity research that language effects are found for the dominant native language, rather than for the second language (Bassetti, 2007; Phillips & Boroditsky, 2003). However, these assessments often relied on participants' self-evaluations of their language dominance and language proficiency. To address this issue, the current study includes proficiency tests for English (language of testing), Ukrainian, and Russian, as well as a self-rated Bilingual Linguistic Profile (BLP, Gertken et al., 2014) to comprehensively assess the proficiency differences.

Typologically, the two languages are linguistic cousins, both belonging to the East Slavic branch of the Indo-European language family (Kortmann & Auwera, 2011), which shares significant historical, lexical, and grammatical similarities. They have a considerable overlap in vocabulary, grammar, and pronunciation characteristics, setting them apart from other Slavonic languages. Various studies indicate that Ukrainian and Russian share about 55%-62% of their vocabulary, a lexical distance akin to that between Portuguese and French (Steinback, 2015). Like other Indo-European languages, Ukrainian and Russian incorporate grammatical gender, categorising nouns as feminine, masculine, or neuter. These languages are highly inflectional with overt gender systems, where gender influences noun declension and adjective endings (Budzhak-Jones, 1997). In Ukrainian, nouns are divided into three genders, with syntactic agreement indicating gender, except for invariably gender-neutral plural nouns (Rusanivskyj et al., 2004). Russian follows a similar division, but with an uneven distribution: 46% of nouns are masculine, 41% feminine, and 13% neuter. The masculine gender, being most prevalent, is often considered the default (Corbett, 1991; 2007). The lack of extensive research on Ukrainian gender distribution leaves the question of whether it follows a similar pattern open.

Although nouns in Ukrainian and Russian neither change according to genders nor have gendered articles, grammatical gender affects the declension of nouns and endings in both languages. In Ukrainian language, masculine gendered animate and inanimate nouns typically have consonant endings (e.g., дім [dim] – house), while feminine gender is predicted by -a / - я endings (e.g., кава [kava] – coffee, історія [istoriia] – history). Most abstract nouns are feminine (Pugh & Press, 1999), regardless of the ending (e.g., радість [radist'] – joy, тиша [tysha] – quiet). Neuter nouns have three possible endings: -o , -e, -ння / -ття (дерево [derevo] – tree, сонце [sontse]– sun, кохання [kokhannia] – love) (Bezpoiasko et al., 1993; Gorpynyč, 2004).

Similarly, in Russian, endings of nouns suggest their grammatical gender: masculine nouns end with a consonant or -й, feminine nouns end with -a or -я, while neuter nouns have -o / -e endings. There is also a large number of exceptions, such as nouns ending with a soft sign -ь, that can refer either to masculine or feminine nouns. In both languages, grammatical gender is semantically and morphologically assigned, affecting adjectives, pronouns, and determiners (Basova et al., 2003), and is a mandatory feature for nouns except in plural forms (Gorpynyč, 2004). The described grammatical gender distribution in Ukrainian and Russian provides a well-suited setting for investigating grammatical gender effects on cognitive processes. It presents an opportunity to go beyond investigating a three-gendered grammatical system, but analysing language effects when 2L1s have contrasting three-gendered systems.

2. Aims and the scope of the current study

This study aims to shed light on what (if any) effects two partially contrasting three-gendered grammatical systems have on cognitive processes of simultaneous bilinguals. While research has examined the impact of single three-gendered systems (Konishi, 1993; Pavlidou & Alvanoudi, 2019; Sera et al., 2002), little is known about the cognitive implications of simultaneously acquiring two languages with differing grammatical features (Bassetti, 2007). We hypothesised that simultaneous bilinguals would demonstrate a language effect similar to that of sequential bilinguals – specifically, they would demonstrate the influence of grammatical gender on categorisation, despite prior research suggesting that gender effects are limited to speakers of two-gendered languages because the binary nature of the system makes grammatical gender more salient (Sera et al., 2002). We expect to observe a grammatical gender effect, by employing a more rigorous stimuli design encompassing grammatical genders both matching and mismatching across languages, coupled with the inclusion of languages where grammatical gender is manifested through diverse grammatical features rather than articles. Additionally, we aim to investigate whether the presence of neuter gender in the stimuli (Experiment 1) would affect the observed grammatical gender effects, compared to Experiment 2, where it was absent. If grammatical gender effects have an online nature, as shown in previous studies (Sato & Athanasopoulos, 2018), we would expect stronger effects in Experiment 2, compared to Experiment 1, as the absence of neuter gender would amplify the contrast between masculine and feminine gender, enhancing the observed effects in the real-time of task completion.

To investigate our hypothesis, we adapted a similarity judgment paradigm where participants rated the similarity of pairs of stimuli, comprising depicted conceptually neutral nouns (e.g., a notebook), presented alongside a picture of a male or female character (e.g., a

ballerina) on a 9-point Likert scale (Phillips & Boroditsky, 2003). The tasks in both experiments were conducted in English (starting with the participant's information sheet in the first email until debriefing). This was done to prevent participant from actively using either of their L1s. The current paradigm was chosen for several reasons. Firstly, it has been used many times, yielding mixed results with speakers of three-gendered languages. However, it has never been used to our knowledge with a three-gendered language omitting the neuter gender as presented in Experiment 2. Using the same task ensures that any effects observed can be attributed to our experimental manipulation rather than any potential confounds of the task itself. Secondly, it was employed due to the high salience of gender/sex in the task (Samuel et al., 2019), laying the groundwork for subsequent exploration of more subtle, implicit effects of gender on cognitive processes.

Experiment 1 aims to provide initial understanding of the grammatical gender effects of Ukrainian and Russian on categorisation, in contrast to English monolingual controls. In the first part of this experiment, we look at the interaction between group (Ukrainian-Russian bilinguals vs English monolinguals) and condition (whether the noun's grammatical gender matches or mismatches the character's biological sex) and whether it had any influence on similarity ratings (Likert scores). Here we anticipate that Ukrainian-Russian bilinguals will show a stronger effects of condition on the similarity ratings compared to English monolinguals. The stimuli include nouns with matching grammatical genders in Ukrainian and Russian (e.g., "pencil"- masculine in both, "candle" – feminine in both, "tree" - neutral in both). Confirming this prediction would reaffirm the original findings by Phillips & Boroditsky (2003) and demonstrate that presence of neuter gender does not negate the language effects. In the second part, when looking at the results of the bilingual group only, we analyse ratings based on participants' most proficient language (Ukrainian or Russian). Stimuli were chosen to include noun-character pairs with contrasting grammatical genders in

Ukrainian and Russian languages (e.g., “a basket” – masculine in Ukrainian, feminine in Russian – paired with a ballerina (female character); “an iron” – masculine in Russian, feminine in Ukrainian – paired with a king (male character)). We predict that bilinguals will rate pairs as more similar when the grammatical gender of the object (masculine or feminine) in their more proficient language is congruent with the character’s biological sex (male or female).

Experiment 2 contains only masculine and feminine nouns, investigating whether excluding neuter gender strengthens the grammatical gender effects. The manipulation here directly addresses a central question in the field regarding the possibility that the presence of neuter gender impairs language effects. The question is whether this happens at a general or a local level. In other words, does the presence of the neuter gender in the grammatical system of a language attenuates effects of gender on categorisation across the board, or are such attenuating effects only observable when the neuter gender is used as part of the similarity judgments that participants are asked to perform. Similar to Experiment 1, we anticipate to find grammatical gender effects on similarity ratings in the Ukrainian-Russian bilingual group but not in the English monolingual group. Within the Ukrainian group, the impact of language proficiency on ratings is also explored.

Overall, we expect to find a significant effect of grammatical gender on categorisation of simultaneous bilinguals, irrespective of the contrasting three-gendered systems of Ukrainian and Russian. The outcomes of this study are expected to highlight the influence grammatical gender has on cognitive processes, shedding more light on how complex and contrasting linguistic systems shape human cognition.

3. Method

Materials and analysis codes can be found on the Open Science Framework (OSF): https://osf.io/3xgaw/?view_only=d061634113d14fa098fb8c2eacb4d81e.

3.1. Experiment 1

Participants. 63 Ukrainian-Russian simultaneous bilingual speakers (with English as a foreign language) and 37 English monolingual speakers completed the study online in exchange for time compensation in a form of a £10 Amazon voucher. After examining their linguistic profiles and responses, 51 Ukrainian-Russian bilinguals (48 females; $Mean_{age}$ = *32,* SD_{age} = *10*) and 24 English monolinguals (9 females; $Mean_{age}$ = *30,* SD_{age} = *13*) were included in the analysis. Exclusion criteria encompassed speaking other gendered languages (n = 22) or consistently selecting a '1' rating on the Likert scale, indicating inattention to instructions or lack of engagement (n = 3). Among the bilingual group, 66.7% (*n* = 34) had a postgraduate degree, 23.5% (*n* = 12) had an undergraduate degree, 2% (*n* = 1) had a college degree, and 7.8% (*n* = 4) had high school education or less. In contrast, among the monolingual group, 50% of participants (*n* = 12) had a postgraduate degree, 25% (*n* = 6) had an undergraduate degree, and 25% of participants (*n* = 6) had a college degree, with no participants having only finished high school.

The bilingual participants proficiency in Ukrainian, Russian, and English was assessed using standardised language tests. For Ukrainian and Russian, advanced ZNO Tests (External Independent Assessment) were used (Ukrainian Centre for Educational Quality Assessment, 2020). These standardised university entrance examinations evaluate participants' language skills up to the C2 proficiency level, thereby mitigating potential ceiling effects of L1 proficiency in our study. English proficiency was determined through the Oxford Quick Placement Test (Oxford University Press, 2001) or existing IELTS

certification (Cambridge University Press, 2021). Acceptable scores were set at 67% for the OQPT and 5.5 for the IELTS, both equivalent to the B2 (Upper-Intermediate) level. ZNO tests classify Ukrainian and Russian proficiency levels between C1 (advanced) and C2 (proficient).

The bilingual participants reported an average age of 8.68 years (SD = 3.21) for acquiring English as a foreign language (L2), with a minimum proficiency level of Upper-Intermediate. The majority of participants demonstrated higher proficiency scores in Ukrainian (57.38%, *n* = 29), as opposed to Russian (22.95%, *n* = 12), or equal proficiency in both (19.67%, *n* = 10). The proficiency scores ranged widely, indicating no ceiling effects (see Table 1).

Table 1

Proficiency Scores and Distribution of Ukrainian-Russian Bilingual Participants in Experiment 1

Language	**Mean Proficiency Score (100 maximum)**	**SD**	**Range**	**Percentage (Number) of Participants**
Ukrainian	65.68	18.39	18.75 - 93.75	57.38% (29)
Russian	59.84	14.90	25.00 - 87.50	22.95% (12)
Equal proficiency in both	57.29	13.55	37.50 - 81.25	19.67% (10)

Participants completed the study online, after being recruited through social media or through posters at [ANONYMISED]. The gender imbalance in bilingual participants, predominantly female, resulted from the data collection occurring after the onset of the war in Ukraine. However, as Flaherty (2001) notes, such a discrepancy in participants' gender is unlikely to significantly affect the responses. Besides, we used separate cumulative link mixed models for each experiment to investgate whether there was an effect of participants' gender (see supplementary materials for full analysis and results). However, the absence of a

significant three-way interaction between group (Ukrainian-Russian bilingual vs English monolingual), participant's gender (male vs female), and grammatical gender (masculine vs feminine vs neuter) suggested that the gender imbalance in the bilingual group did not appear to disproportionately affect the main findings of the study.

Materials

Pre-test. A pre-test was conducted to select conceptually gender-neutral items for the main experiment, following the approach of Sato & Athanasopoulos (2018). Ten Ukrainian-Russian-English speakers (5 females; *Mean age = 26, SD age = 4*) and ten English monolinguals (4 females; *Mean age = 31, SD age = 10*) were recruited. None of the participants took part in the main study. Participants were shown 137 black-and-white object images one by one and asked to rate each picture on a 7-point Likert scale ranging from "very feminine" (1) to "very masculine" (7). The objects were divided into five groups based on their grammatical genders in Ukrainian and Russian: (1) 20 nouns masculine in Russian and feminine in Ukrainian, (2) 24 nouns feminine in Russian and masculine in Ukrainian, (3) 31 nouns feminine in both languages, (4) 31 nouns masculine in both languages, and (5) 31 nouns neutral in both languages. All images, presented against a greyscale and white background to avoid colour biases, were sourced from the Bank of Standardised Stimuli (Brodeur et al., 2014).

The pre-test yielded 50 conceptually neutral items (*Mean = 4.01; SD = 0.13*), which were then divided into the five categories (see Table 2): (1) nouns with masculine grammatical gender in both Russian and Ukrainian languages, (2) feminine grammatical gender in both Russian and Ukrainian, (3) feminine in Russian, masculine in Ukrainian, (4) feminine in Ukrainian and masculine in Russian, and (5) neutral in both. A slight imbalance between stimuli (3) and (4) is not anticipated to impact our results, as they will be analysed

collectively. This will yield a total of 20 nouns with matching grammatical gender in both languages, 20 nouns with mismatching grammatical gender, and 10 neuter fillers.

Table 2

Example of stimuli used for both Experiment 1 and 2

Type of stimuli	Example (Russian)	Example (Ukrainian)	English Translation	Number of Items
Masculine in both Russian and Ukrainian	миндаль (mindal)	мигдаль (myhdal)	almond	10
Feminine in both Russian and Ukrainian	свечка (svechka)	свічка (svichka)	candle	10
Feminine in Russian, Masculine in Ukrainian	лодка (lodka)	човен (choven)	boat	8
Feminine in Ukrainian, Masculine in Russian	муравей (muravei)	мураха (murakha)	ant	12
Neutral in both languages (Experiment 1 only)	яблоко (yabloko)	яблуко (yabluko)	apple	10

Main testing. In the main experiment, participants were presented with a hundred pairs, each consisting of one of the 50 selected conceptually neutral unlabelled black-and-white objects and one of the16 characters: 8 female images (a queen, a bride, a witch, a smurf, a ballerina, a girl, a pensioner, an ogre) and 8 male images (a king, a groom, a giant, a smurf, an architect, a boy, a man, an ogre). Each depicted noun was presented once with a male character and once with a female character, resulting in 100 pairs. Pairs were presented in a randomised order. Each participant had to provide a similarity rating on the Likert scale from 1 (not similar) to 9 (very similar) with each pair displaying the object on the left and the character on the right of the screen.

Procedure and design

To conduct the experiment, we utilised the Gorilla Experiment Builder software. Upon registration, participants received an introductory email containing the participant

information sheet and a link to the experiment. After signing a consent form, they were redirected to the main task, which they accessed on their personal laptops or computers.

Both groups undertook the same experimental task in English. The instructions were similar to those from Phillips and Boroditsky (2003, p. 929): “In this study, you will see pairs of pictures appear on the screen. In each pair, there will be a picture of a person on the left and a picture of an object or animal on the right. You will see a scale where 1 = not similar and 9 = very similar. For each pair of pictures, please choose a number between 1 and 9 to indicate how similar you think the two pictures are. Try to use the whole scale (give some 1’s and some 9’s and some of all the numbers in-between). Please respond with the first answer that comes to mind”.

Each object-person pair remained on the screen until participants selected “Next”. Once they moved on to the next pair, they could not change their answer. After completing the task, participants were asked what criteria were used to rate the pairs to determine whether they detected the experiment’s aim and used grammatical gender as a task-solving strategy. None of the participants reported reliance on grammatical gender or language in general. Instead, responses were reported to be influenced by associations with films or cartoons, shapes, or random guesses. Ukrainian-Russian bilingual participants then completed a Bilingual Language Profile (BLP, Gertken et al., 2014) questionnaire and two proficiency tests (Oxford University Press, 2001; Ukrainian Center for Educational Quality Assessment, 2020). The monolingual group only completed the BLP to identify any gendered language knowledge potentially affecting results. Additionally, we monitored the real-time completion of the experiment. In those instances where participants substantially exceeded the expected average response times or stopped during the task, their participation was manually excluded (6 bilingual and 9 monolingual participants), given the importance of capturing responses on the first-impression basis.

Analysis

For each experiment, data analysis involved cumulative link mixed models in RStudio (version 2022.07.22, R Core Team, 2022), using the ordinal package (Christensen, 2019), with similarity ratings as the dependent variable. Previous study that replicatd the original experiment by Phillips and Boroditsky (2003) employed linear mixed-effects models (Elpers et al., 2022), highlighting their advantages, such as incorporating both fixed and random effects and analyzing non-averaged data (Baayen et al., 2008; Vasishth & Broe, 2011). However, as the analysis includes Likert scale and ordinal data, we used cumulative link mixed models (CLMMs) instead. Similarly to linear mixed effects models, CLMMs also accommodate multiple sources of error variance as random variables, such as participant variability and the gender of depicted characters (Bross, 2019). Yet, CLMMs are more suited for analysing ordinal data, as they account for possibility of varying distances between levels of the rating scale (Ackerman, 2018).

We divided the analysis into two parts. The first part involved a comparative analysis of responses from both Ukrainian-Russian bilingual and English monolingual participants. We focused on how the interaction between grammatical gender congruence of the pairs (grammatical gender of the object was congruent or incongruent in both Russian and Ukrainian with the biological sex of the character) and the participant group (Ukrainian-Russian or English) influenced the Likert scores. The maximal model that converged included random intercepts for participants and items. The detailed analysis is available on OSF (https://osf.io/3xgaw/?view_only=d061634113d14fa098fb8c2eacb4d81e).

Secondly, to investigate deeper the effects of two contrasting three-gendered languages, we conducted an analysis comparing Ukrainian-Russian bilinguals only, based on their most proficient language. In the current study, we approached bilingualism as a continuum and measured it as a continuous variable by subtracting Russian proficiency from

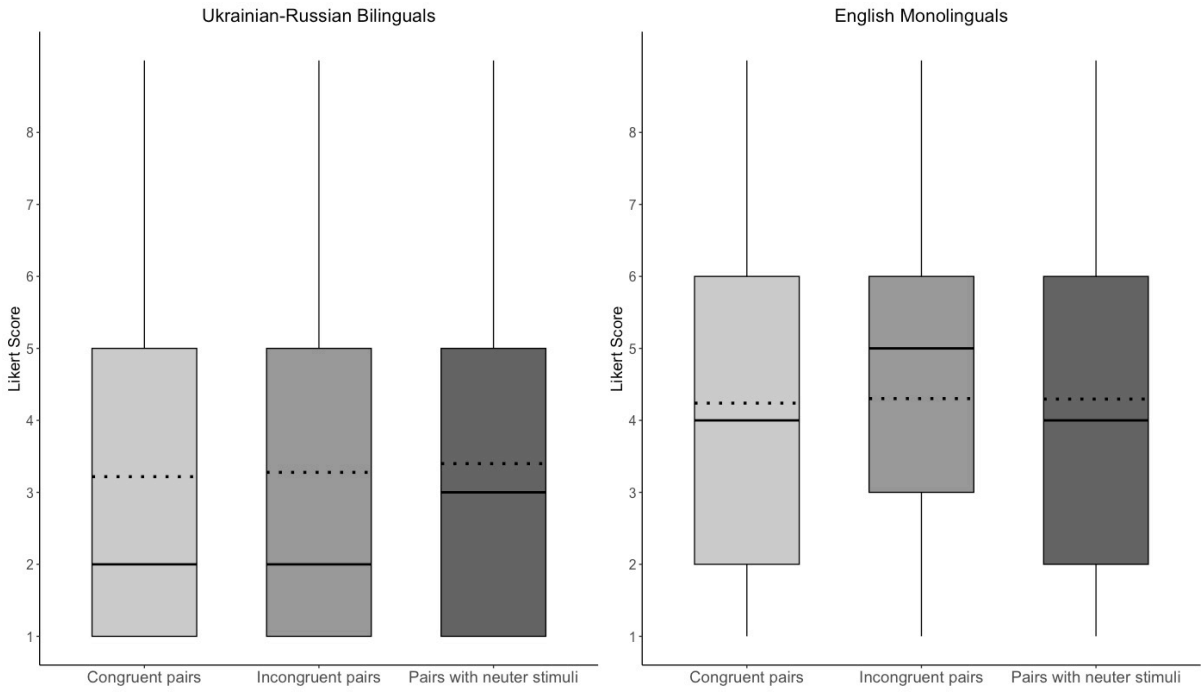
Ukrainian proficiency scores, resulting with the scale -100 being only proficient in Russian and +100 only proficient in Ukrainian. Participants with equal proficiency scores were included in the analysis with the coefficient score 0. Here, we examined how the congruence of an object's grammatical gender in L1 with the character's biological sex (and its incongruence in 2L1) interacted with language proficiency to affect similarity ratings. A maximal model in this part also included random intercepts for both participants and items.

Results

Comparing the Ukrainian-Russian bilingual and English monolingual participants. In this analysis, we included stimuli where the grammatical gender of nouns was either congruent or incongruent with the character's biological sex in both Ukrainian and Russian. An example of this would be "a ballerina" (female) and "a pen" (feminine in both Ukrainian and Russian) or "a ballerina" and "an almond" (masculine in both). Our expectation was that Ukrainian-Russian bilinguals would show stronger grammatical gender effects compared to English monolinguals. Specifically, we predicted that congruent pairs, where the character's biological sex is congruent with the object label's grammatical gender in 2L1s, would receive higher similarity ratings. For instance, Ukrainian-Russian bilinguals were anticipated to rate a congruent pair, such as "a ballerina" and "a pen", as more similar than incongruent pairs like "a king" and "a pen". English monolinguals were not expected to show any significant trends.

Figure 1

Comparison of Likert scores across conditions for Ukrainian-Russian Bilinguals and English Monolinguals: mean (dotted line) and median (solid line) differences in congruent, incongruent, and neuter stimuli pairs in Experiment 1



Comparing the mean responses of Ukrainian-Russian bilingual participants in the congruent (*Mean = 3.22, SD = 2.41*) and incongruent (*Mean = 3.28, SD = 2.41*) pairs revealed nearly identical ratings, contrary to our predictions (see Figure 1). Notably, bilingual participants displayed slightly higher, but not statistically significant, average responses for stimuli with neuter grammatical gender (*Mean = 4.00, SD = 2.55*). In contrast, English monolingual participants consistently assigned similar ratings across all conditions (congruent: Mean = 4.24, SD = 2.38; incongruent: Mean = 4.30, SD = 2.17; neuter: Mean = 4.30, SD = 2.38), indicating that condition type did not notably influence their judgments of object-character similarity.

We built a cumulative link mixed model (clmm) to compare two groups of participants looking at the interaction between the group (Ukrainian-Russian bilinguals vs English monolinguals) and pair congruency in both L1s (congruent vs incongruent vs neutral), as a predictor for similarity ratings (Likert scores). Random intercepts were included for participants and items to account for variations specific to each.

The results revealed a statistically significant Group effect, with Ukrainian-Russian bilinguals exhibited lower similarity ratings compared to the English controls (*SE = 0.3318, z = -2.771, p = 0.006*). However, there were no statistically significant main effects for pair congruency (*SE = 0.2194, z = 0.165, p = 0.869)* or for the interaction between the two variables. Specifically, the lack of significant group - condition interaction (*SE = 0.1376, z = 0.888, p = 0.3744*) demonstrated that, in contrast to our hypothesis, Ukrainian-Russian bilinguals did not rate incongruent pairs as less similar compared to the congruent pairs.

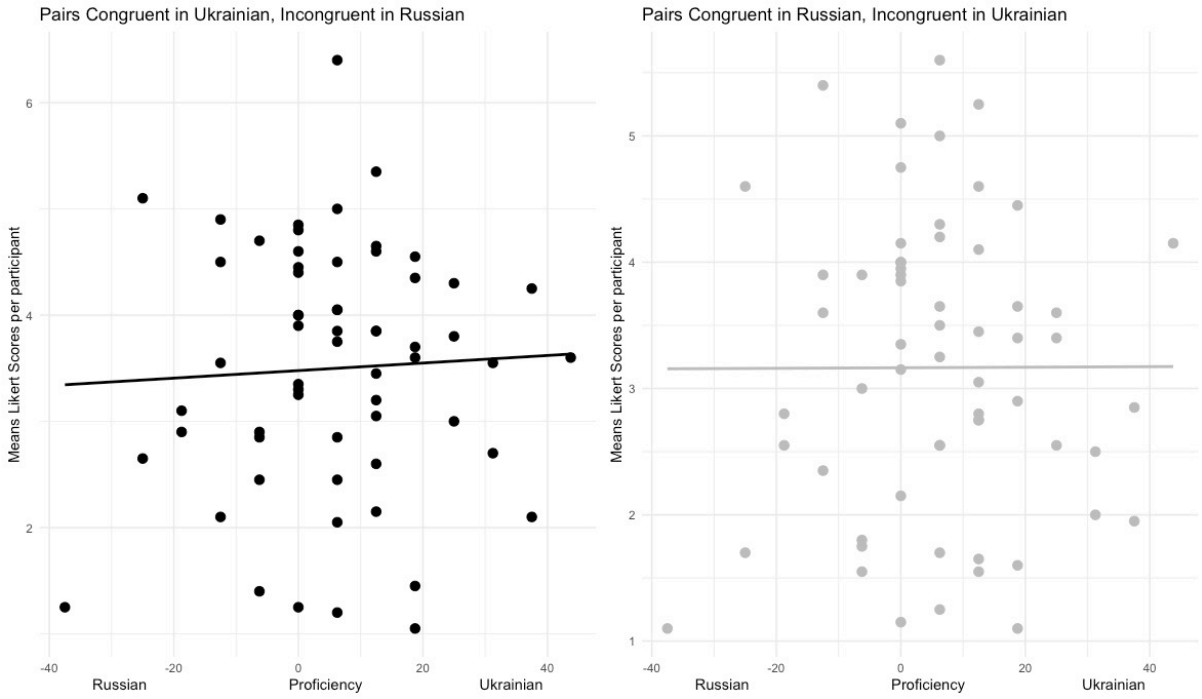
Comparing Ukrainian-Russian simultaneous bilinguals based on the Language proficiency in L1 and 2L1. To compare the results of Ukrainian-Russian bilinguals only and the investigate the effect of the more proficient first language (L1 or 2L1) on similarity ratings, we conducted a separate analysis with different stimuli. This included noun pairs where grammatical gender matched the character's biological sex in one language but not the other. For example, “a queen” and “an onion” (masculine in Russian, feminine in Ukrainian) were congruent in Ukrainian but incongruent in Russian. Conversely, “a king” and “a sock” (feminine in Ukrainian, masculine in Russian) were congruent in Russian and incongruent in Ukrainian.

Ukrainian-Russian bilinguals assigned ratings to pairs congruent in Ukrainian *(Mean = 3.50, Range = 2.92 - 4.08)* and pairs congruent in Russian *(Mean = 3.17, Range = 2.59 -*

3.74) when their proficiency was higher in Ukrainian (Figure 2). However, the differences in ratings were minimal and statistically non-significant, against our expectations.

Figure 2

Mean Likert Scale Responses from Experiment 1 per participant (Ukrainian-Russian bilinguals only) by Language Proficiency for pairs of stimuli where characters' biological sex and objects' grammatical gender are (a) congruent in Ukrainian and incongruent in Russian, and (b) congruent in Russian and incongruent in Ukrainian



In the second cumulative link mixed model, we explored whether Likert scores were influenced by the interaction between condition (biological sex and grammatical gender congruent in Ukrainian and incongruent in Russian vs congruent in Russian and incongruent in Ukrainian) and language proficiency (-100 to 100, with -100 being only Proficient in Russian, to 100 –only proficient in Ukrainian). The maximum convergence model included

random intercepts for participants and items to account for participant-specific and item-specific variations. Contrary to our predictions, we found no significant effects for the condition-proficiency interaction (*SE = 0.005, z = -0.784, p = 0.433*), demonstrating that bilingual participants with higher proficiency in Russian did not assign higher ratings to the pairs that were congruent in Russian and incongruent in Ukrainian. Furthermore, no significant main effects for condition (*SE = 0.3241, z = -0.741, p = 0.459*) or language proficiency (*SE = 0.0104, z = 0.725, p = 0.468)* were found. Overall, our findings for the stimuli with mismatching grammatical gender in 2L1s suggest that neither the individual variables nor their interaction significantly contributed to participants' similarity ratings.

In summary, Experiment 1 revealed that gender congruence of noun-character pairs had no statistically significant impact on similarity ratings. Moreover, an unexpected pattern emerged, as Ukrainian-Russian bilinguals consistently rated objects as less similar than their English monolingual counterparts across all conditions. Our findings in this experiment align with the claims by Sera et al. (2002) that the presence of neuter grammatical gender may negate grammatical gender effects in speakers of three-gendered languages.

3.2. Experiment 2

The results from Experiment 1 suggest that including a neutral gender may have mitigated the significance of the language effect by diminishing the salience of grammatical gender. This raised the possibility that excluding neutral gender from the study design could affect the findings, particularly if the grammatical gender effects are online in nature and arise from real-time language effects. Therefore, in this study, we largely retained the methodology used in Experiment 1 but excluded the neuter gender from the stimuli.

Participants

40 English monolinguals and 70 Ukrainian-Russian bilinguals were recruited. After analysing their linguistic profile and responses, 64 bilinguals (44 females; *Mean age = 30, SD age = 12*) and 34 monolinguals (18 females; *Mean age = 26, SD age = 6*) were included in the analysis. Exclusions were due to participants either knowing other gendered languages (n = 6) or consistently using a single value on the Likert scale (n = 6), suggesting a potential lack of engagement or failure to follow instructions. The demographic distribution of the bilingual group in Experiment 2 was consistent with that of Experiment 1. As in the previous experiment, the largest proportion of bilingual participants held postgraduate degrees: 42.2% (*n* = 27). This was followed by 31.3% (*n* = 20) with undergraduate degrees, 18.8% (n = 12) with a high school diploma, and 7.8% (*n* = 5s) with a college degree. For the monolingual group, the distribution shifted slightly from Experiment 1. While postgraduate degrees remained the most common (35.3%, *n* = 12), the proportions for college and undergraduate degrees changed. In Experiment 1, college and undergraduate diplomas were equally represented, but in Experiment 2, 32.4% (*n* = 11) had a college diploma, 23.5% (*n* = 8) held an undergraduate degree, and 8.8% (*n* = 3) had a high school education. Similarly to Experiment 1, no effects of participants' gender on their ratings were found (see Tables 4 and 5 in supplementary materials).

Analogously to the first experiment, we assessed bilingual participants' linguistic profiles and proficiency of Ukrainian, Russian, and English. Participants were recruited online and via posters at [ANOMYMISED]. The bilingual participants reported acquiring English (L2) at an average age of 9 years (*Range = 4-20*) and had at least an upper-intermediate proficiency level. Among them, 72% of participants demonstrated higher proficiency scores in Ukrainian and 28% in Russian. None of the participants reported using

grammatical gender as a conscious strategy. The proficiency scores varied widely (see Table 3), demonstrating that ceiling effects were absent.

Table 3

Proficiency Scores of Ukrainian-Russian bilinguals in Experiment 2

Language	Mean Proficiency Score (100 maximum)	SD	Range	Percentage (Number) of Participants
Ukrainian	65.2	19.3	12.5 – 100	72% (46)
Russian	51.7	12.7	25 – 81.2	17% (11)
Equal proficiency in both	55.4	8.41	43.8 - 68.8	11% (7)

Materials

As with Experiment 1, participants were asked to rate object-character pairs using a 1 (not similar) to 9 (very similar) Likert scale. The stimuli consisted of 40 conceptually neutral black-and-white objects, categorised as follows: 10 masculine in both Russian and Ukrainian, 10 feminine in both languages, 8 feminine in Russian but masculine in Ukrainian, and 12 feminine in Ukrainian but masculine in Russian. In addition, 16 characters (8 male, 8 female; the same as in Experiment 1) were used. To compensate for the reduction in stimuli due to the exclusion of neutral grammatical gender, we adjusted the number of trials in this experiment. Specifically, we paired each object with every character (rather than just one male and one female pairing per item as in Experiment 1), resulting in 640 unique pairs. This adjustment was made for two main reasons. First, the exclusion of neuter gender reduced the overall number of stimuli, which could have impacted the statistical power of the study, while increasing the number of trials helped to counterbalance this reduction. Second, in Experiment 1, pairings were pseudorandomised to minimise the risk of semantic associations (e.g., avoiding obvious pairings like “a broom” with “a witch”). In Experiment 2, to eliminate

this potential confound entirely, each object was paired with every character, thus increasing variability and reducing the chance of unintended semantic associations. The trial order was randomised for each participant, with objects presented on the left and characters on the right of the screen.

To ensure the validity of the data, we adopted enhanced measures, including comprehensive guidelines detailing the necessary procedures and environment for successful task completion. Additionally, participants were observed during the experiment. Any participant observed becoming distracted or communicating in their native languages was excluded from the analysis (13 bilingual and 11 monolingual speakers).

Procedure and design

The approach for Experiment 2 closely followed that of Experiment 1, but with the inclusion of participant observation conducted via Zoom. An experimenter monitored each session to ensure that participants were focused, free from distractions, and not using their native language during the task. All interactions were done in English and if participants needed clarifications, they did so in English as well. In Experiment 2, we also modified the verbal instructions to emphasise the use of the entire response scale (1 to 9). This adjustment was made based on observations from Experiment 1, where some participants tended to limit their responses to a narrower range of the scale. The experimenter used intonation to explicitly highlight this request during the verbal instructions, while maintaining the original instructions from Experiment 1. The modified instructions, given in English, were as follows: “In this study, you will see pairs of pictures appear on the screen. In each pair, there will be a picture of a person on the left and a picture of an object or animal on the right. You will see a scale where 1 = not similar and 9 = very similar. For each pair of pictures, please choose a number between 1 and 9 to indicate how similar you think the two pictures are. Try to use the

WHOLE scale (give some 1's and some 9's and some of all the numbers in-between). Please respond with the first answer that comes to mind. Please try not to be distracted and avoid communicating with anyone (unless necessary) until the experiment is complete." The final sentence, instructing participants to avoid distractions and communication, was added specifically for experiment 2 to help maintain task focus.

The analytical approach remained consistent with that of experiment 1, employing a similar structure for the cumulative link mixed models. The analysis comprised two parts. In the first part, we compared the responses of English monolinguals and Ukrainian-Russian bilinguals. This comparative analysis explored the effects of pair congruence (congruent vs incongruent in both Russian and Ukrainian) and group (Ukrainian-Russian bilinguals vs English monolinguals) interaction on Likert scores. The second part focused on the examining responses from Ukrainian-Russian bilinguals only, assessing the effect of pair congruence (congruent in Ukrainian / incongruent in Russian vs congruent in Russian / incongruent in Ukrainian) and language proficiency (-100 to 100, with -100 being only Proficient in Russian, to 100 –only proficient in Ukrainian) interaction on similarity ratings. In both parts of the analysis, the maximum convergence models included random intercepts for participants and items.

Results

Comparing the Ukrainian-Russian bilingual and English monolingual participants.

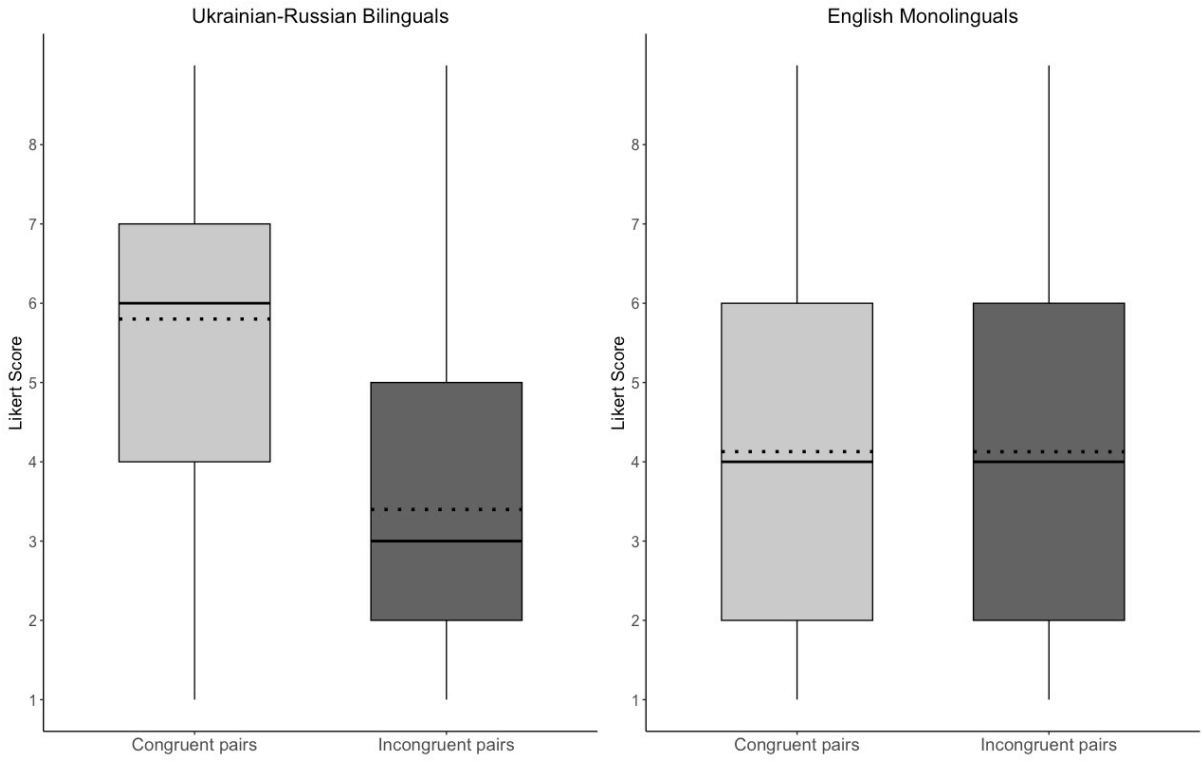
Consistent with our predictions, bilinguals assigned significantly higher ratings to pairs with congruent biological sex and grammatical gender in both L1 and 2L1 (*Mean = 5.8, SD = 2.0*), as opposed to the incongruent pairs (*Mean = 3.4, SD = 1.73*). Besides, as confirmed by pairwise comparison, bilinguals rated congruent pairs significantly higher than monolingual participants (*Mean = 4.12, SD = 2.4*). As for the incongruent pairs (fig.3), Ukrainian-Russian

bilinguals tended to rate them significantly lower (*Mean = 3.4, SD = 1.74*) than English controls (Mean = 4.13, SD = 2.4). For the English monolingual group, there was no significant difference between the 'congruent' and 'incongruent' conditions (estimate = 0.0115, SE = 0.0245, z = 0.47, p = 0.639).

Figure 3

Comparison of Likert scores across conditions for Ukrainian-Russian Bilinguals and English Monolinguals: mean (dotted line) and median (solid line) differences in congruent and incongruent stimuli pairs in Experiment 2.

** p < .05. ** p < .01. *** p < .001.*



Analogously to the first experiment, a cumulative link mixed model examined the interaction between the proup (bilingual vs. monolingual) and condition (congruent vs. incongruent in both L1s), as a predictor for similarity ratings. The results revealed a

statistically significant group effect for Ukrainian-Russian bilinguals (*SE = 0.0888, z = 16.38, p < 0.001*). We also found significant effects for the bilingual group - condition interaction, indicating that bilinguals assigned significantly lower rating to the incongruent pairs (*SE = -1.9301, z = -55.15, p < 0.001*) than English monolinguals. These findings confirmed our hypothesis that matching grammatical gender in both languages of bilinguals significantly affects their categorisation once neutral gender is excluded from the testing conditions.

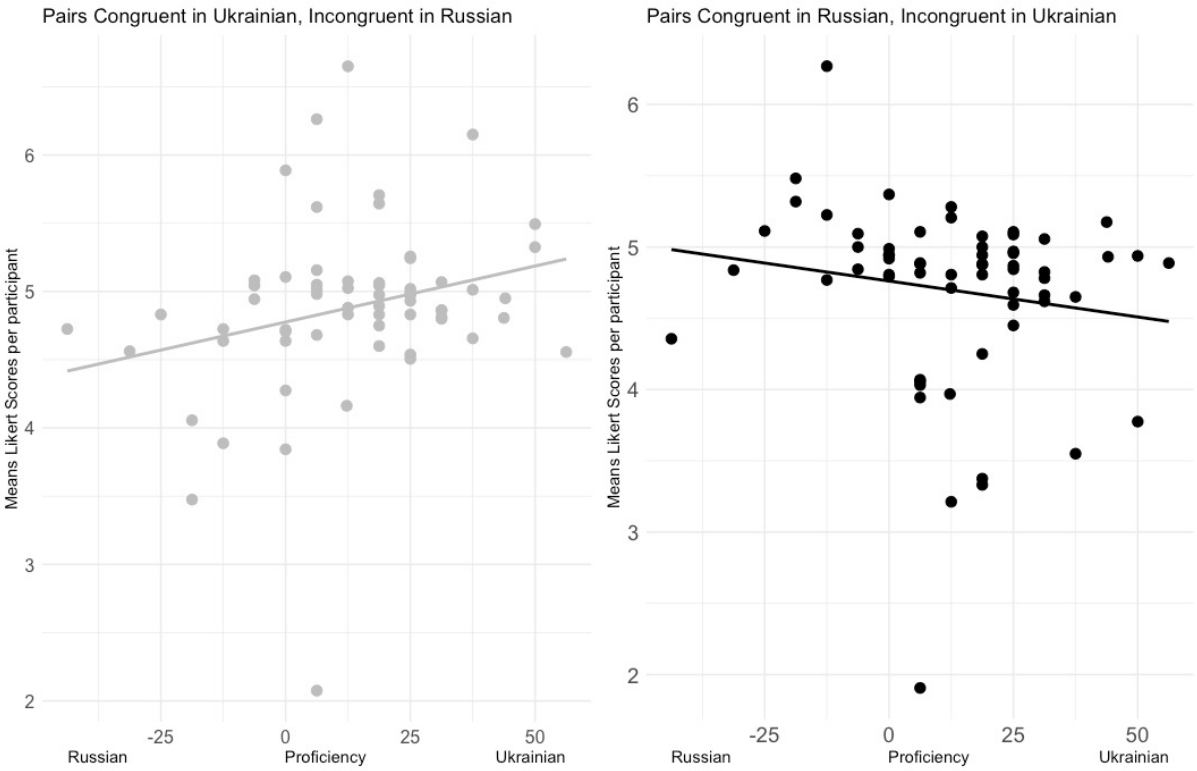
Comparing Ukrainian-Russian simultaneous bilinguals based on the language proficiency in L1 and 2L1. Figure 4 illustrates a clear difference in ratings, in line with our expectations. Ukrainian-Russian bilingual participants who were more proficient in the Ukrainian language gave significantly higher similarity ratings to object-character pairs where the object's grammatical gender in Ukrainian was congruent to the character's biological sex (*Mean = 4.99, SD = 2.26*), compared to pairs congruent in Russian (*Mean = 4.56, SD = 2.38*). Conversely, those with higher proficiency in Russian tended to give significantly higher ratings to pairs congruent in Russian (*Mean = 5.12, SD = 2.24*) than to incongruent ones (*Mean = 4.54, SD = 2.26*).

The designed cumulative link mixed model tested the impact of the interaction between condition (congruent with Ukrainian language and incongruent with Russian vs congruent with Russian language and incongruent with Ukrainian) and language proficiency (-100 to 100). While no significant main effect for Condition (*SE = 0.0292, z = 0.475, p = 0.635*), a significant main effect of Proficiency (*SE = 0.0022, z = -1.960, p = 0.05*) was observed. Besides, as predicted, a significant interaction was found between condition and proficiency (*SE = 0.0013, z = 8.622, p < 0.001*). This suggests that the interaction between most proficient L1 of a simultaneous bilingual and condition had a significant impact on

categorisation, and those bilingual participants that were more proficient in Ukrainian rated pairs that were congruent in Ukrainian and incongruent in Russian as more similar, and vice versa for those more proficient in Russian.

Figure 4

Mean Likert Scale Responses from Experiment 2 per participant (Ukrainian-Russian bilinguals only) by Language Proficiency for pairs of stimuli where characters' biological sex and objects' grammatical gender are (a) congruent in Ukrainian and incongruent in Russian, and (b) congruent in Russian and incongruent in Ukrainian



4. Discussion

The current study aimed to explore how language, grammatical gender in particular, affects cognitive processes of Ukrainian-Russian bilinguals in an all-English context. The group was chosen for several reasons. First, Ukrainian and Russian grammatical systems have nouns with both matching and contrasting grammatical gender across languages. Secondly, both languages have three grammatical genders (masculine, feminine, and neuter). Incorporating Ukrainian and Russian languages is beneficial for linguistic relativity research because, unlike previously studied languages such as Italian, Spanish, French, or German, they lack articles that could conflict with the biological sex of the referent. Instead, grammatical gender in Ukrainian and Russian is predominantly marked through noun, adjective, and sometimes verb endings. This distinct morphosyntactic feature — where gender is conveyed directly through morphological changes rather than through articles or fixed gender markers — has often been overlooked in existing research focused on languages with different gender-marking strategies.

Moreover, one of our research interests in the present study was to contribute to the discussion of whether gender effects arise online or offline, by examining whether having neuter gender embedded in the grammatical systems of both Ukrainian and Russian would lead to diminished grammatical gender effects. Therefore, we adapted one of the seminal studies on grammatical gender (Phillips & Boroditsky, 2003), while manipulating grammatical gender in Ukrainian and Russian, as well as presence (Experiment 1) and absence (Experiment 2) of neuter gender in testing conditions.

In Experiment 1, we observed a lack of significant effects of grammatical gender and group, as well as their interaction, when comparing the ratings of bilingual and monolingual participants. Additionally, we found no effects of the interaction between language

proficiency and grammatical gender in Ukrainian-Russian bilinguals, indicating that their more proficient language had little to no effect on similarity judgements. Such findings align with previous research that reported lack of grammatical gender effects on speakers of three-gendered languages, such as German (Sera et al., 2002; Vigliocco et al., 2005).

However, after excluding neuter gender in Experiment 2, a significant interaction between group and condition was found when comparing bilingual and monolingual groups, indicating that Ukrainian-Russian bilinguals rated higher those pairs where grammatical gender of an object in both Ukrainian and Russian was congruent with biological sex of a character, compared to the incongruent pairs. Additionally, a significant interaction between condition and language proficiency was observed, when only simultaneous bilinguals' results were analysed. The latter demonstrated that bilinguals with higher proficiency in Ukrainian rated those pairs as more similar where grammatical gender and biological sex were congruent in Ukrainian and incongruent in Russian. The analoguous effect was observed for speakers more proficient in Russian, as they perceived the pairs congruent in Russian to be more similar than those congruent in Ukrainian.

Before discussing differences between the experiments, we should first explore the possible reasons for the null results in Experiment 1. The absence of significant results in the first experiment might be attributed to several factors. Firstly, as suggested by Sera et al. (2002) and Vigliocco, et al. (2005), three-gendered grammatical systems may not show effects as strong as those in two-gendered languages with more direct and intuitive associations between grammatical gender and natural gender, which can lead to stronger perceptual biases. In contrast, three-gendered systems which include a neuter gender, introduce a level of grammatical complexity that may obscure the relationship between gender and categorisation. The neuter gender, in particular, could have reduced the salience of masculine and feminine distinctions, thereby weakening potential gender effects.

Secondly, the broader lack of support for findings using this paradigm may reflect ongoing issues related to the replication crisis in linguistic relativity research. As mentioned earlier, most previous attempts to replicate Phillips and Boroditsky (2003) have not yielded significant results, except for Pavlidou and Alvanoudi (2019). For instance, Elpers et al. (2022), even with an increased sample size, failed to provide the significant results using the linear mixed effects models, though analysis using the t-tests showed significance. This issue is exacerbated by methodological variations and by the use of different statistical analyses across studies that employ the same paradigm, which makes it challenging to compare results consistently. Finally, the unique linguistic profiles of participants, which often differ across research contexts, add another layer of complexity. Previous studies that used a similarity judgement task also focused on bilingual participants, but there is limited consistency in how those participants were selected or their linguistic profiles were characterised. Key details, such as whether participants spoke other gendered languages and the criteria used for proficiency self-assessment are often not reported in sufficient detail. This variability makes it difficult to draw meaningful comparisons across studies, as differences in participant characteristics could significantly influence the observed effects – or the lack thereof - of grammatical gender. However, a key unifying factor between our study and those conducted by Sera et al. (2002) and Vigliocco et al. (2005) is the inclusion of neuter gender in the stimuli. This suggests that the presence of neuter gender may have influenced the absence of grammatical gender effects observed across these studies.

The discrepancy in language effects between Experiments 1 and 2 could be attributed to variations in experimental design, such as increased number of stimuli, variation in instructions or participant observations in Experiment 2, as well as lack of neuter gender in the task. While we initially hypothesised that the absence of neuter gender would be primarily driving the observed differences, it is important to consider that other

methodological changes may also have contributed. First, the increased the number of stimuli in Experiment 2 likely enhanced statistical power, providing a clearer picture of language effects that might have been less detectable in Experiment 1. Besides, increased number of pairs allowed us to account for the possible semantic associations in Experiment 2 that could have emerged in Experiment 1 (e.g., pairing ‘a broom and ‘a witch’ together). To examine the potential outcomes of using only the stimuli from Experiment 1 within the context of Experiment 2, an additional analysis was conducted with this subset. This analysis, which included 72 pairs of stimuli from Experiment 1, confirmed a robust and significant effect for both types of stimuli, consistent with the results obtained from the full stimuli set in Experiment 2. These findings strengthen the interpretation that the absence of neuter-gender stimuli in Experiment 2 may be a driving factor behind the observed grammatical gender effects, further validating our findings. Detailed analysis have been included in the supplementary materials (pp. 9–10). Second, the modified verbal instructions emphasised the use of the entire scale (1 to 9), which may have influenced participants to use a broader range of responses. Third, the addition of participant observation via Zoom allowed the experimenter to ensure that participants remained focused and did not revert to their native language.

However, it is also possible that the observed differences in the results were primarily due to the absence of the neuter grammatical gender, as hypothesised. This effect may be explained by considering the distinction between online and offline language processing discussed in the literature. According to both the label-feedback (Lupyan, 2020) and structural-feedback hypotheses (Sato & Athanasopoulos, 2018), online effects occur when language actively modulates perception and decision-making in real time, influenced by top-down feedback from specific linguistic labels and broader structural patterns respectivelly. In Experiment 1, the inclusion of neuter gender may have diluted the salience of masculine and

feminine categories, reducing the immediate impact of gender cues on participants' judgments. Neuter nouns might have introduced a neutral, less distinctive category that disrupted the online processing of gender, as it did not align with the binary masculine-feminine distinction. This aligns with findings from previous research, which suggest that the presence of a third, neuter category can weaken the perceptual link between grammatical and natural gender - not in the offline manner as claimed by Sera et al. (2002), but during the process of task completion. In Experiment 2, by excluding neuter gender, the task environment emphasised over the course of the experiment the binary masculine-feminine distinction, creating a feedback loop where the structure of the gender system becomes more entrenched and influences real-time (online) processing more strongly. Without the neutral baseline provided by neuter nouns, participants were more inclined to use the salient gendered cues actively, resulting in more pronounced effects. This suggests that the grammatical gender effects observed in Experiment 2 were primarily driven by the immediate, context-sensitive use of gender information (i.e., online effects), but also by the reinforcing influence of the underlying linguistic structure on cognitive processing (i.e., a structural feedback effect).

In sum, our study shows that such an effect does not have its roots in the mere presence of the neuter gender in a languages's grammatical system, but rather arises online, as a function of the absence of the neuter gender in the task. Such an interpretation is compatible with modern accounts of the mechanisms underpinning linguistic relativity effects, such as the label-feedback hypothesis (Lupyan, 2012) and the structural-feedback hypothesis (Sato & Athanasopoulos, 2018). These findings also align with earlier research and demonstrate that three-gendered languages do indeed impact cognitive processes, such as categorisation. Furthermore, the language effects are present even when grammatical genders

do not match in the two languages of simultaneous bilinguals, as they rely on the grammatical gender of their more proficient language.

The complexity of our findings underscores the necessity for more nuanced research methodologies. The similarity judgment task is merely the first step in analysing gender effects within our new group of participants. We suggest that future research employ more rigorous methodologies to further investigate these effects. For instance, incorporating neurophysiological measures, such as event-related potentials (ERPs), to better elucidate the effects of grammatical gender on bilingual cognition. This could be done by adapting previously used paradigms by Sato, et. al. (2020) or Boutonnet, et al. (2012) to investigate whether grammatical gender primes conceptual or semantic represention (looking at N300 or Left Anterior Negativity respectively) in speakers of three-gendered compared to speakers of two-gendered speakers that were used in these two studies. Additionally, we recommend expanding the range of stimuli used to test speakers of multiple three-gendered languages. For example, future research could include nouns that have masculine or feminine grammatical gender in one language (L1) and neuter gender in the second language (2L1). This expansion would provide further insights into the influence of grammatical gender on bilingual cognition, grammatical gender representation in simultaneous/early bilingual's mind, and contribute to the broader field of linguistic relativity.

References

Ackerman, L. (2018). *Dealing with ordinal data*. https://lmackerman.com/notebooks/ordinal_data.nb.html#what_is_%E2%80%9Cordinal%E2%80%9D_data

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*(1), 89–96. https://doi.org/10.1017/S1366728905002397

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

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*(2), 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., & Bylund, E. (2023). Cognitive restructuring: Psychophysical measurement of time perception in bilinguals. *Bilingualism: Language and Cognition*, *26*(4), 809–818. https://doi.org/10.1017/S1366728922000876

Athanasopoulos, P., Bylund, E., & Casasanto, D. (2016). Introduction to the Special Issue: New and Interdisciplinary Approaches to Linguistic Relativity: New Approaches to Linguistic Relativity. *Language Learning*, *66*, 482–486. https://doi.org/10.1111/lang.12196

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

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

Bassetti, B., & Filipović, L. (2022). Researching language and cognition in bilinguals. *International Journal of Bilingualism*, *26*(1), 3–12. https://doi.org/10.1177/13670069211022860

Belacchi, C., & Cubelli, R. (2012). Implicit knowledge of grammatical gender in preschool children. *Journal of Psycholinguistic Research*, *41*(4), 295–310. https://doi.org/10.1007/s10936-011-9194-y

Bender, A., Beller, S., & Klauer, K. C. (2011). Grammatical Gender in German: A Case for Linguistic Relativity? *Quarterly Journal of Experimental Psychology*, *64*(9), 1821–1835. https://doi.org/10.1080/17470218.2011.582128

Bezpoiasko, O. K., Gorodenska, K. G., & Rusanivskiy, V. M. (1993). *Grammar of Ukrainian language. Morphology* (M. S. Tymoshyk & L. L. Ščerbatenko, Eds.; 1st ed.). Lybid.

Bilaniuk, L., & Melnyk, S. (2008). A Tense and Shifting Balance: Bilingualism and Education in Ukraine. *International Journal of Bilingual Education and Bilingualism*, *11*, 340–372. https://doi.org/10.1080/13670050802148731

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Boroditsky, L., & Schmidt, L. A. (2000). Sex, Syntax, and Semantics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *22*(22). https://escholarship.org/uc/item/0jt9w8zf

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

Bross, F. (2019). Using Mixed Effect Models to Analyze Acceptability Rating Data in Linguistics. Version 1.0. *Mimeo*. https://doi.org/www.fabianbross.de/mixedmodels.pdf.

Budzhak-Jones, S. (1997). Quantitative analysis of gender assignment in mono/bilingual discourse. *Journal of Quantitative Linguistics*, *4*(1–3), 67–91. https://doi.org/10.1080/09296179708590080

Cambridge University Press. (2021). *IELTS (International English Language Testing System)* | *Cambridge English*. https://www.cambridgeenglish.org/exams-and-tests/ielts/

Cao, Y., Majid, A., & Vanek, N. (2024). Not all verbal labels grease the wheels of odor categories. PsyArXiv. https://doi.org/10.31234/osf.io/g873s

Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, *66*(3), 714–730. https://doi.org/10.1111/lang.12192

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulu, O., Pita, R., & Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26).

Christensen, R. H. B. (2019). *Cumulative Link Models for Ordinal Regression with the R Package ordinal* (Version R package version 2019.12-10) [Computer software]. http://www.cran.r-project.org/package=ordinal/

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, *10*(2), 137–152. https://doi.org/10.1177/13670069060100020201

Cook, V. J. (2002). Background to the L2 user. In V. J. Cook (Ed.), *Portraits of the L2 user* (pp. 21–28). Multilingual Matters. https://www.degruyter.com/document/doi/10.21832/9781853595851/pdf?licenseType=restricted#page=9

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181–200. https://doi.org/10.1017/S1366728903001123

Csernicskó, I., & Máté, R. (2017). *Bilingualism in Ukraine: Value or Challenge? 10*, 14–35. http://dx.doi.org/10.1515/sm-2017-0001

Eberhard, D., Simons, G., & Fennig, C. (2019). *Ethnologue: Languages of the World, 22nd Edition*.

Elpers, N., Jensen, G., & Holmes, K. J. (2022). Does grammatical gender affect object concepts? Registered replication of Phillips and Boroditsky (2003). *Journal of Memory and Language*, *127*, 104357. https://doi.org/10.1016/j.jml.2022.104357

Flaherty, M. (2001). How a Language Gender System Creeps into Perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31. https://doi.org/10.1177/0022022101032001005

Gentner, D., & Goldin-Meadow, S. (2003). *Language in Mind: Advances in the Study of Language and Thought*. The MIT Press. https://doi.org/10.7551/mitpress/4117.001.0001

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual language profile. In *Measuring L2 Proficiency: Perspectives from SLA* (pp. 208–225).

Gorpynyč, V. O. (2004). *Morfolohiya ukrayins'koyi movy: Pidruchnyk dlya studentiv vyshchykh navchal'nykh zakladiv [Morphology of the Ukrainian language: A textbook for students of higher educational institutions]* (1st ed.). Academy.

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, *104*(2), 417–426. https://doi.org/10.1016/j.cognition.2006.07.008

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1), 9–28. https://doi.org/10.1515/probus-2016-0002

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

Kortmann, B., & Auwera, J. V. D. (2011). *The Languages and Linguistics of Europe: A Comprehensive Guide*. De Gruyter, Inc. http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=765858

Kulyk, V. (2023, January 7). *Мова та ідентичність в Україні на кінець 2022-го [Language and identity in Ukraine by the end of 2022]*. Zbruč. https://zbruc.eu/node/114247

Lucy, J. A. (1992). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis* (Vol. 13, Issue 13, p. xi 328). Cambridge University Press.

Lucy, J. A. (1997). Linguistic Relativity. *Annual Review of Anthropology*, *26*(Volume 26, 1997), 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*. https://doi.org/10.3389/fpsyg.2012.00054

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, *171*, 172–179. https://doi.org/10.1016/j.cognition.2017.10.010

Mitrofanova, N., Rodina, Y., Urek, O., & Westergaard, M. (2018). Bilinguals' sensitivity to grammatical gender cues in Russian: The role of cumulative input, proficiency, and dominance. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01894

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

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*(6), 405–428. https://doi.org/10.1080/00207594.2012.743665

Pavlidou, T.-S., & Alvanoudi, A. (2013). Grammatical Gender and Cognition. *Versita/de Gruyter*, *2*(2), 109–124.

Pavlidou, T.-S., & Alvanoudi, A. (2019). Conceptualizing the world as 'female' or 'male': Further remarks on grammatical gender and speakers' cognition. *Selected Papers on Theoretical and Applied Linguistics*, *23*(0), Article 0. https://doi.org/10.26262/istal.v23i0.7351

Phillips, W., & Boroditsky, L. (2003). Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25). https://escholarship.org/uc/item/31t455gf

Poftak, N., & Shykula, D. (2022). The Ongoing Language Discussion in Ukraine. In *Being Ukraine: An Introduction to Europe's Eastern Vanguard*. Connecticut College. https://openpress.digital.conncoll.edu/beingukraine/chapter/chapter-7/

Pugh, S., & Press, I. (1999). *Ukrainian: A Comprehensive Grammar*. Routlege.

R Core Team. (2022). *R: The R Project for Statistical Computing*. The R Project for Statistical Computing. https://www.r-project.org/

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

Rusanivskyj, V. M., Taranenko, O. O., & Zjabljuk, M. P. (2004). *Encyclopaedia of the Ukrainian language* (M. P. Bažan, Ed.; 2nd ed.). Kyiv: 'Ukrainian Encyclopaedia'.

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

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

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), 377–397. https://doi.org/10.1037/0096-3445.131.3.377

Shumlianskyi, S. (2010). *Conflicting abstractions: Language groups in language politics in Ukraine*. *2010*(201), 135–161. https://doi.org/10.1515/ijsl.2010.007

Speed, L. J., & Majid, A. (2019). Linguistic features of fragrances: The role of grammatical gender and gender associations. *Attention, Perception, & Psychophysics*, *81*(6), 2063–2077. https://doi.org/10.3758/s13414-019-01729-0

Steinback, S. (2015). *Lexical-distance-among-the-languages-of-europe-2-1-mid-size.png 1,853×1,642 pixels*. Lexical Distance among the Languages of Europe. https://upload.wikimedia.org/wikipedia/commons/f/f3/Lexical-distance-among-the-languages-of-europe-2-1-mid-size.png

Ukrainian center for educational quality assessment. (2020). *Ukrainian center for educational quality assessment*. https://testportal.gov.ua/en/

Vanek, N., Sóskuthy, M., & Majid, A. (2021). Consistent verbal labels promote odor category learning. *Cognition*, *206*, 104485. https://doi.org/10.1016/j.cognition.2020.104485

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), 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Whorf, B. L. (1956). *Language, Thought, and Reality, second edition: Selected Writings of Benjamin Lee Whorf* (Caroll, J.B., Ed.). 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*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Supplementary materials

1. The role of participants' gender on the similarity ratings

To address the potential effects of gender imbalance in the bilingual participant groups in both Experiment 1 and 2 (see table 1), we have conducted a separate analysis examining gender as a potential predictor on similarity ratings. Except for "female" and "male", our participants also had choices "prefer not to say" and "other", however, none of our participants chose either of these options.

Table 1.

Number of participants in Experiments 1 and 2, based on their gender

	Experiment 1		Experiment 2	
	Ukrainian-Russian bilinguals	English monolinguals	Ukrainian-Russian bilinguals	English monolinguals
Male	3	15	20	16
Female	48	9	44	18
Total number of participants	51	24	64	34

In Experiment 1, we designed two cumulative link mixed models for each part of the analysis. In the first part where we compare the similarity ratings between Ukrainian-Russian bilinguals and English monolinguals, the maximum convergence model included a three-way interaction between *group* (Ukrainian-Russian vs English), *participant's gender* (male vs female) and *grammatical gender of an item in both L1s* (feminine vs masculine vs neuter) and whether there were any effects on Likert scores. By-item and by-participant random intercepts were also added to the model. The results of the clmm model are presented in table 2.

Table 2.

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 1, Comparing Ukrainian-Russian Bilinguals and English Monolinguals, including Participant's Gender as a predictor

	Estimate	Std. Error	z-value	Pr(>|z|)
Group (Ukrainian)	**-1.349**	**0.392**	**-3.445**	**0.001*** **
Participant's gender (male)	-0.659	0.560	-1.178	0.239
Grammatical gender of an item in both languages (masculine)	-0.141	0.236	-0.598	0.550
Grammatical gender of an item in both languages (neuter)	-0.087	0.236	-0.369	0.712
Group (Ukrainian) * Participant's gender (male)	**1.605**	**0.704**	**2.279**	**0.023***
Group (Ukrainian) * Grammatical gender of an item in both languages (masculine)	0.187	0.167	1.120	0.263
Group (Ukrainian) * Grammatical gender of an item in both languages (neuter)	0.205	0.168	1.226	0.220
Participant's gender (male)*Grammatical gender of an item in both languages (masculine)	**0.486**	**0.236**	**2.064**	**0.039***
Participant's gender (male)*Grammatical gender of an item in both languages (neuter)	0.303	0.237	1.282	0.200
Group (Ukrainian) * Participant's gender (male)*Grammatical gender of an item in both languages (masculine)	-0.450	0.299	-1.505	0.132
Group (Ukrainian) * Participant's gender (male)*Grammatical gender of an item in both languages (neuter)	-0.175	0.302	-0.579	0.562

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

If there were an effect of participant gender specifically within the bilingual group, we would expect a significant three-way interaction between **group** (Ukrainian-Russian vs. English), **participant's gender** (male vs. female), and **grammatical gender of the item in both L1s** (masculine vs. feminine). This interaction would indicate that male Ukrainian-Russian bilingual participants rated items with masculine grammatical gender significantly higher than those with feminine grammatical gender. However, this three-way interaction was not significant *(SE = 0.302, z = -0.579, p = 0.562),* suggesting that participant gender did not have a unique effect within the bilingual group based on grammatical gender.

However, the only significant fixed effect was **group (Ukrainian-Russian bilinguals),** which was unrelated to the prediction about participant gender effects. In addition, two interactions reached significance. Firstly, **group and participant's gender significant** interaction *(SE = 0.704, z = 2.279, p = 0.023)* indicated that, on average, male Ukrainian-Russian bilingual participants rated all items higher than female bilingual participants. Secondly, significant interaction between **participant's gender and grammatical gender in both L1s** *(SE = 0.236, z = 2.064, p = 0.039)* suggested that male participants from both groups (Ukrainian-Russian bilinguals and English monolinguals) rated items with masculine grammatical gender significantly higher than those with feminine grammatical gender. However, while male participants tended to rate masculine-gendered items higher overall, this pattern was consistent across both groups (including English monolinguals that were not aware of grammatical gender systems in Ukrainian and Russian), with no evidence of a unique gender effect specific to the bilingual group.

In the second part of the analysis, we examined a model with a four-way interaction involving **condition** (congruent in Ukrainian vs. congruent in Russian), **grammatical gender in Ukrainian** (masculine vs. feminine), **participant's gender** (male vs. female), and **group proficiency** (ranging from -100 for full proficiency in Russian to +100 for full proficiency in Ukrainian). A significant four-way interaction would indicate that male Ukrainian-Russian bilingual participants with higher proficiency in Ukrainian rated stimuli with masculine grammatical gender in Ukrainian higher than those with feminine grammatical gender. And vice versa for female participants. However, no significant interactions confirming this prediction was found (table 3).

Table 3.

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 1, Comparing Ukrainian-Russian bilingual group based on Language Proficiency, including Participant's Gender as a predictor

	Estimate	**Std. Error**	**z-value**	**Pr(>|z|)**
Condition (congruent in Russian)	-0.369	0.438	-0.842	0.4
Participant's gender (male)	0.731	0.549	1.331	0.183
Grammatical gender of an item in Ukrainian (masculine)	-0.280	0.462	-0.607	0.544
Language Proficiency	0.007	0.011	0.59	0.555
Condition (congruent in Russian) * Participant's gender (male)	0.036	0.319	0.113	0.91
Condition (congruent in Russian) * Grammatical gender of an item in Ukrainian (masculine)	0.316	0.653	0.484	0.628
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine)	0.169	0.332	0.507	0.612
Condition (congruent in Russian) * Language Proficiency	-0.004	0.007	-0.568	0.57
Participant's gender (male) * Language Proficiency	0.006	0.038	0.165	0.869
Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	-0.003	0.008	-0.427	0.669
Condition (congruent in Russian) * Participant's gender (male) * Language Proficiency	-0.071	0.468	-0.152	0.879
Condition (congruent in Russian) * Participant's gender (male) * Language Proficiency	-0.022	0.022	-1.008	0.313
Condition (congruent in Russian) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	0.004	0.011	0.356	0.722
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	-0.019	0.023	-0.823	0.411
Condition (congruent in Russian) * Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	0.024	0.032	0.726	0.468

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

As can be seen in table 3, no significant effects were found for either interaction/ main effect, suggesting that participant's gender did not significantly impact their responses when rating nouns that have mismatching grammatical gender in Ukrainian and Russian.

To check if participants' gender affected their ratings in Experiment 2, we build clmm models analogously to Experiment 1. In the first part of the analysis where we analysed ratings for nouns with matching grammatical gender across Ukrainian and Russian (table 4), the maximum convergence model included a four-way interaction between Condition (congruent in both L1s vs incongruent in both L1s), Grammatical gender of an item in both languages (masculine vs feminine), Participant's gender (male vs female), Group (Ukrainian-Russian bilinguals vs English monolinguals). It also included random intercepts for participants and items.

Table 4.

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian Bilinguals and English Monolinguals, including Participant's Gender as a predictor

	Estimate	Std. Error	z value	Pr(>|z|)
Condition (incongruent in both L1s)	-0.051	0.048	-1.057	0.291
Grammatical gender of an item in both languages (masculine)	-0.098	0.059	-1.669	0.095
Participant's gender (male)	**0.736**	**0.236**	**3.121**	**0.002****
Group (Ukrainian-Russian bilinguals)	**1.760**	**0.193**	**9.142**	**< 0.001*****
Condition (incongruent in both L1s) * Grammatical gender of an item in both languages (masculine)	0.080	0.068	1.181	0.237
Condition (incongruent in both L1s) * Participant's gender (male)	0.020	0.070	0.287	0.774
Grammatical gender of an item in both languages (masculine)* Participant's gender (male)	0.022	0.070	0.309	0.757
Condition (incongruent in both L1s) * Group (Ukrainian-Russian bilinguals)	**-1.897**	**0.063**	**-29.997**	**< 0.001*****
Grammatical gender of an item in both languages (masculine)* Group (Ukrainian-Russian bilinguals)	0.122	0.063	1.949	0.051
Participant's gender (male)* Group (Ukrainian-Russian bilinguals)	**-0.789**	**0.304**	**-2.597**	**0.009****
Condition (incongruent in both L1s) * Grammatical gender of an item in both languages (masculine) * Participant's gender (male)	-0.044	0.099	-0.442	0.658
Condition (incongruent in both L1s) * Grammatical gender of an item in both languages (masculine) * Group (Ukrainian-Russian bilinguals)	-0.029	0.089	-0.333	0.739
Condition (incongruent in both L1s) * Participant's gender (male) * Group (Ukrainian-Russian bilinguals)	-0.114	0.102	-1.113	0.266
Grammatical gender of an item in both languages (masculine) * Participant's gender (male)* Group (Ukrainian-Russian bilinguals)	-0.021	0.102	-0.202	0.840
Condition (incongruent in both L1s) * Participant's gender (male) * Group (Ukrainian-Russian bilinguals) * Grammatical gender of an item in both languages (masculine)	0.099	0.145	0.683	0.495

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

We did not find any significant interactions either for participant's gender – grammatical gender in 2 L1s (*SE = 0.070, z = 0.309, p = 0.757*), or for the four-way condition-participant's gender – group – grammatical gender interaction (*SE = 0.145, z = 0.683, p = 0.495*), indicating that participants' gender did not have a significant effect on ratings of items with masculine grammatical gender.

In the part 2, the maximum convergence model **condition** (congruent in Ukrainian vs. congruent in Russian), **grammatical gender in Ukrainian** (masculine vs. feminine), **participant's gender** (male vs. female), and **language proficiency** (ranging from -100 for full proficiency in Russian to +100 for full proficiency in Ukrainian). Similarly, to the second clmm model in Experiment 1, no significant effects were found for either interaction/ main effect (table 5), suggesting that participant's gender did not significantly impact their responses when rating nouns that have mismatching grammatical gender in Ukrainian and Russian.

Overall, results from both experiments indicate that participants' gender did not significantly impact ratings or impact the main findings. The gender imbalance in the bilingual groups did not seem to introduce systematic bias into the results, and rating patterns were consistent across male and female participants.

Table 5.

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian bilingual group based on Language Proficiency, including Participant's Gender as a predictor

	Estimate	Std. Error	z-value	Pr(>|z|)
Condition (congruent in Ukrainian)	-0.369	0.438	-0.842	0.4
Grammatical gender of an item in Ukrainian (masculine)				
Participant's gender (male)	0.731	0.549	1.331	0.183
Language Proficiency	0.007	0.011	0.59	0.555
Condition (congruent in Ukrainian) * Grammatical gender of an item in Ukrainian (masculine)	0.036	0.319	0.113	0.91
Condition (congruent in Russian) * Participant's gender (male)	0.316	0.653	0.484	0.628
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine)	0.169	0.332	0.507	0.612
Condition (congruent in Ukrainian) * Language Proficiency	-0.004	0.007	-0.568	0.57
Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency				
Participant's gender (male) * Language Proficiency	0.006	0.038	0.165	0.869
Condition (congruent in Ukrainian) * Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine)	-0.071	0.468	-0.152	0.879
Condition (congruent in Ukrainian) * Grammatical gender of an item in Ukrainian (masculine)* Language Proficiency	-0.022	0.022	-1.008	0.313
Condition (congruent in Ukrainian) * Participant's gender (male) * Language Proficiency	0.004	0.011	0.356	0.722
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	-0.019	0.023	-0.823	0.411
Condition (congruent in Ukrainian) * Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	0.024	0.032	0.726	0.468

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

2. Reanalysis of Experiment 2 results using stimuli from Experiment 1 (excluding neuter-gender stimuli)

In this subsection, we examine whether significant results would still emerge for the second experiment when using only a subset of stimuli from Experiment 1. Experiment 1 included 50 conceptually neutral items, 10 of which had neuter grammatical gender in both Ukrainian and Russian. After excluding the neuter stimuli, the subset consisted of 40 items, each paired once with a male and once with a female character (object-character pairs the same as in the Experiment 1). This yielded a total of 80 pairs chosen from 640 pairs.

The current analysis was divided into two parts, following the structure of the analyses in both Experiments 1 and 2: (1) the analysis of stimuli with matching grammatical gender in both L1s and (2) the analysis of stimuli with mismatching grammatical gender in the two L1s.

For the first part, we focused on the similarity ratings of Ukrainian-Russian bilinguals and English monolinguals, considering stimuli with matching grammatical gender in both Ukrainian and Russian. A cumulative link mixed model (CLMM) was employed, identical to the model used in prior analyses, to test whether the interaction between Group (Ukrainian-Russian bilinguals vs. English monolinguals) and Condition (congruent in both Ukrainian and Russian vs. incongruent in both Ukrainian and Russian) had a significant effect on similarity ratings (Likert scores). The maximum convergence model also included random intercepts for participants and items.

The results (table 5) revealed a robust statistically significant effect of the interaction (*SE = 0.105, z = -16.620, p <0.001 for Ukrainian-Russian group and incongruent condition*), consistent with the findings from the full analysis in Experiment 2. These results indicate that, even when using only the stimuli from Experiment 1 without adding additional pairs to mitigate potential semantic associations, Ukrainian-Russian bilinguals rated incongruent pairs in both of their L1s as significantly less similar compared to congruent pairs.

Table 5.

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian Bilinguals and English Monolinguals, using pairs of stimuli from Experiment 1 (without neuter gender)

	Estimate	**Std. Error**	**z-value**	**Pr(>|z|)**
Group (Ukrainian-Russian bilinguals)	**1.291**	**0.165**	**7.806**	**<0.001*** **
Condition (incongruent in 2L1s)	0.075	0.074	1.008	0.313
Group (Ukrainian-Russian bilinguals) * Condition (incongruent in 2L1s)	**-1.750**	**0.105**	**-16.620**	**<0.001*** **

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Similarly, in the second part of the analysis, we focused on the similarity ratings of pairs containing items with mismatching grammatical gender in the two L1s, as rated by Ukrainian-Russian bilinguals. To do so, we developed a clmm model to examine the interaction between language proficiency (ranging from -100 for exclusive proficiency in Russian to +100 for exclusive proficiency in Ukrainian) and condition (congruent in Ukrainian vs. congruent in Russian). By-item and by-participants random intercepts were also included in the model.

The results (Table 6) revealed a statistically significant interaction between proficiency and condition (*SE = 0.004, z =2.684, p = 0.007 for pairs congruent in Ukrainian*), also consistent with the findings from the full analysis in Experiment 2. This suggests that participants with higher proficiency in Ukrainian rated pairs congruent in Ukrainian (but incongruent in Russian) as more similar, whereas participants with higher proficiency in Russian showed the opposite pattern, rating pairs congruent in Russian as more similar.

Table 6.

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian Bilinguals based on the Language Proficiency, using pairs of stimuli from Experiment 1

	Estimate	**Std. Error**	**z-value**	**Pr(>|z|)**
Language Proficiency	-0.004	0.003	-1.260	0.208
Condition (congruent in Ukrainian)	0.101	0.087	1.155	0.248
Language Proficiency * Condition (congruent in Ukrainian)	**0.010**	**0.004**	**2.684**	**0.007****

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Overall, this analysis confirms that even when using pairs from Experiment 1 without expanding the stimuli to include pairings with all eight characters (as opposed to just two), significant effects of language on bilinguals' categorisation are still observed for the subset of data from the second experiment.

3. Distribution of Likert scores in Experiments 1 and 2

To explore the potential effects of the modifications in instructions between experiment 1 and 2, we analysed the differences in distributions of Likert scores. While the main body of instructions remained unchanged, in Experiment 2 participants received the instruction verbally in addition to seeing it on the screen, with the experimenter emphasising the need to use the whole range of scores from 1 to 9.

3.1. Comparing Ukrainian-Russian bilinguals and English monolinguals (using stimuli with matching grammatical gender in both L1s)

First, we examined the histograms of Likert scores for the two groups of participants using pairs that contain stimuli that had matching grammatical gender. In Experiment 1, the histograms for the two groups (English and Ukrainian-Russian) show distinct patterns (fig. 1). For instance, the English monolingual group displays a fairly uniform distribution across the Likert scale with no clear peaks. The responses are spread quite evenly, although there is a slight increase in frequency around the middle scores (4-6). On the other hand, the Ukrainian-Russian bilingual group shows a different pattern, with a noticeable concentration of responses at the lower end of the Likert scale (1-3), that was also reflected in figure 1 in the manuscript. The distribution is positively skewed (table 7), indicating that participants from this group tended to select lower scores more frequently. In Experiment 2, the histograms illustrate a shift in the response patterns for both groups. The English group exhibits a more left-skewed distribution compared to Experiment 1, with a higher frequency of responses at the lower end (1-4). This change is supported by an increase in skewness (*from 0.056 in Experiment 1 to 0.391 in Experiment 2*) and a slight increase in variance (*5.737*) and standard deviation (*2.395*), as shown in table 7.

Conversely, the Ukrainian-Russian group shows a more balanced distribution with a peak around the middle of the Likert scale (scores 4-6). The responses appear less skewed than in Experiment 1 (*from 0.847 in Experiment 1 to 0.177 in Experiment 2*), indicating a broader spread of

scores and a more symmetric pattern, potentially due to the emphasis in the modified instructions. Additionally, the group's variance (4.941) and standard deviation (2.223) were lower than in Experiment 1, suggesting a more consistent use of the scale. The interquartile range (IQR) also narrowed from 4 to 3, reflecting a more concentrated central tendency.

Figure 1.

Distribution of Likert scores in Experiments 1 and 2, by participant group and using stimuli with matching grammatical gender in 2L1s

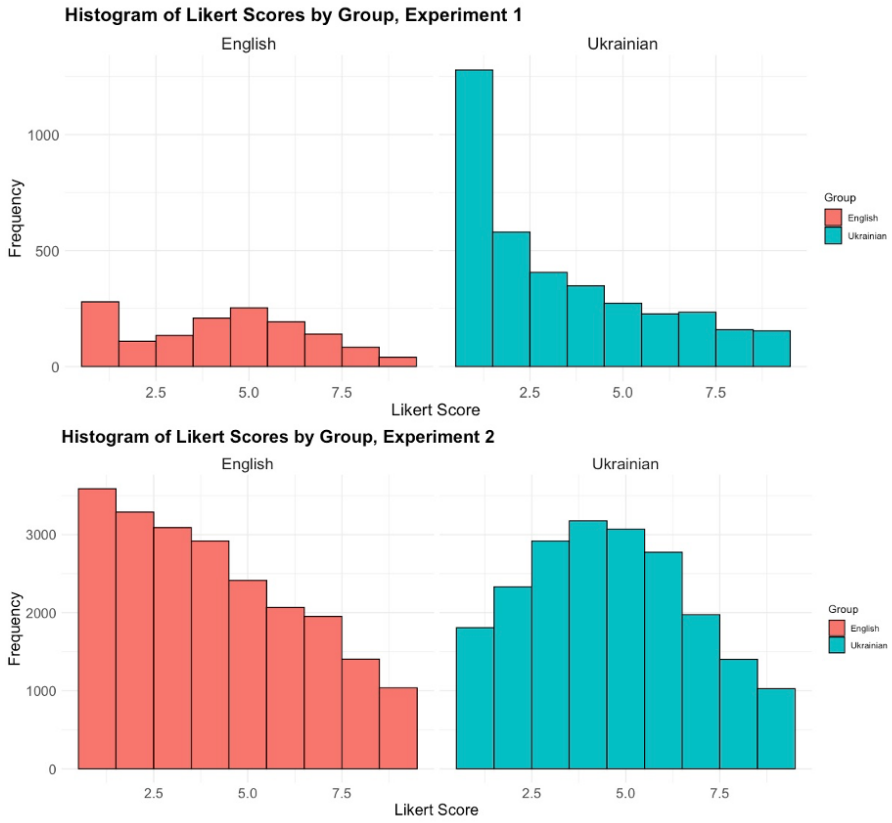


Table 7.

Descriptive statistics for Likert scores across Experiments 1 and 2 by participant group

Variable	Scores for Experiment 1		Scores for Experiment 2	
	English	Ukrainian-Russian	English	Ukrainian-Russian
Range	1-9	1-9	1-9	1-9
Interquartile Range (IQR)	4	4	4	3
Variance	5.269	6.044	5.737	4.941
Standard Deviation	2.295	2.459	2.395	2.223
Skewness	0.056	0.847	0.391	0.177
Kurtosis	2.023	2.516	2.060	2.157

3.2. Results of Ukrainian-Russian bilinguals (using stimuli with mismatching grammatical gender in both L1s)

The distribution of Likert scores for Ukrainian-Russian bilinguals responding to stimuli with mismatching grammatical gender also displayed notable shifts between Experiment 1 and Experiment 2 (figure 2).

In **Experiment 1**, the histogram shows a pronounced skew toward the lower end of the Likert scale, with the majority of responses concentrated between scores 1 and 3. The descriptive statistics (table 8) further support this observation, with a positive skewness of ***0.846***, reflecting the asymmetry of the distribution. The **variance** (*6.347*) and **standard deviation** (*2.519*) highlight substantial variability in the scores, though the distribution is less spread out than in Experiment 2. The kurtosis value of ***2.468*** suggests a heavier tail compared to a normal distribution, indicating some extremity in responses.

Conversely, in **Experiment 2**, the histogram illustrates a more balanced distribution, with a peak around the middle of the scale (scores 4–6). This indicates a broader use of the Likert scale, likely influenced by the emphasis in the modified instructions to use the full range of scores. The descriptive statistics (table 8) show a reduction in **skewness** to ***0.072***,

reflecting a more symmetric response pattern compared to Experiment 1. With the decrease of variance and standard deviation (*to **5.338** and to **2.311 respectively)***, less variability was observed in participant responses. The **kurtosis** also decreased to ***2.018***, indicating a less peaked and more evenly distributed set of responses.

Overall, these results suggest that the modification in instructions had a significant effect on response behaviour, encouraging participants to utilise the entire Likert scale more evenly.

Figure 2.

Distribution of Likert scores of Ukrainian-Russian bilinguals in Experiments 1 and 2, using stimuli with mismatching grammatical gender in 2L1s

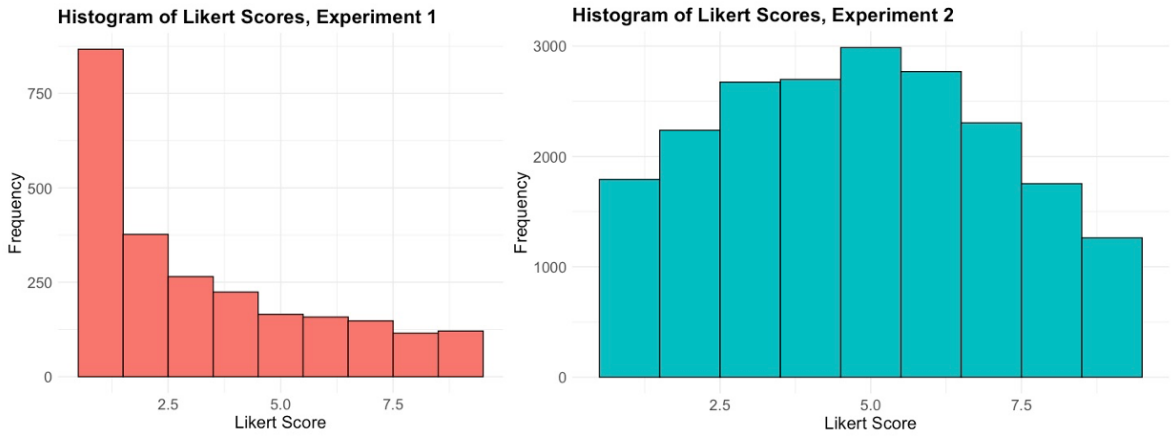


Table 7.

Descriptive statistics for Likert scores across Experiments 1 and 2 for Ukrainian-Russian bilingual group

Variable	Experiment 1	Experiment 2
Range	1-9	1-9
Interquartile Range (IQR)	4	4
Variance	6.347	5.338
Standard Deviation	2.519	2.311
Skewness	0.846	0.072
Kurtosis	2.468	2.018