“She says, he says”: Does the sex of an instructor interact with the grammatical gender of targets in a perspective-taking task?

Steven Samuel, Karen Roehr-Brackin & Debi Roberson

University of Essex

Keywords: bilingualism, grammatical gender, perspective-taking, masculine, feminine, language and thought.

Correspondence concerning this article should be addressed to Debi Roberson, Department of Psychology, Wivenhoe Park, University of Essex, Colchester, CO4 3SQ, UK. Telephone: +44 (0)1206 873710; email address: robedd@essex.ac.uk

Word count = 5836

Abstract

The present study investigated effects of facilitation or interference on object selection by speakers whose L1 marks grammatical gender, even when selecting objects in their L2. Participants were asked to select objects by an instructor, whose biological sex (and voice) was either congruent or incongruent with the grammatical gender of the object to be selected. Bilinguals were expected to find it easier to take alternative perspectives (and to switch perspectives between their own and another's) than monolinguals, due to their proposed superior ability to inhibit prepotent responses. When tested in English, bilinguals whose L1 marked grammatical gender showed no effect of gender congruency in this task, nor did they outperform monolinguals in taking other perspectives or in perspective switching. The present findings are interpreted as evidence that the effects of L1 grammatical gender on tasks performed in an L2 are limited.

Introduction

Almost 60 years since Whorf's observations on the nature of the relationship between natural language and thought (Whorf, 1956) the topic continues to be extensively studied (e.g. Goddard & Wierzbicka, in press; Slobin, 2003), and two main branches of investigation have emerged. One has compared widely differing cultures and languages, such as traditional hunter-gatherer, nomadic or farming communities with native English speakers (Heider & Olivier, 1972; Roberson, Davies & Davidoff, 2000; Roberson, Davidoff, Davies & Shapiro, 2004, 2005) while another has compared cultures with similar levels of urbanisation, education, media exposure and lifestyle, whose languages differ in a critical variable – grammatical gender (e.g. Sera et al., 2002; Vigliocco et al. 2005; Costa et al, 2003; Segel & Boroditsky, 2011; Cubelli, Paolieri & Lotto, 2011; Bender, Beller & Klauer, 2011; Bassetti, 2007). The latter studies may have the advantage over the former in that differences in performance are less likely to result from other factors that might affect an individual's ability to carry out a task, such as cultural environment.

A number of recent studies have compared speakers of a language that does not mark grammatical gender, such as English or Japanese, with speakers of a language that does (e.g. German, French, Spanish or Italian) (e.g. Ramos & Roberson, 2010; Athanasopoulos et al, 2011; Saalbach et al., et al., 2012; Philips & Boroditsky, 2003). In those studies participants were asked either to make judgements of similarity or attribution of masculine or feminine characteristics to inanimate objects or to animals. In the case of inanimate objects, the assignment of grammatical gender to object labels appears largely arbitrary (Corbett, 1991). In the case of animals (e.g.

giraffes in Italian) the generic term is used for both males and females[1] so the assignment of masculine or feminine gender to the generic term also appears somewhat arbitrary. In both cases, investigators have sought to establish whether this apparently arbitrary grammatical gender assignment 'rubs off' on speakers' conceptual representation of the object so that speakers come to perceive objects that do not have biological sex (male or female) to have either masculine or feminine characteristics.

Support for a pervasive effect of grammatical gender comes from a picture-word interference task. Paolieri et al (2011) found that monolingual Italian speakers took longer to name pictures when the grammatical gender of the distractor word and the picture to be named were congruent than when they mismatched, leading the researchers to conclude that the distractor word was processed faster than the picture, resulting in inhibition of the picture name and longer response latencies.

Other researchers have asked bilingual individuals whose first language (L1) carried grammatical gender but whose second language (L2) did not, to make judgments in their L2, reasoning that any predictable attribution of gendered characteristics would indicate a pervasive effect of grammatical gender on semantic representations. For example, Philips and Boroditsky (2003) asked Spanish-English and German-English bilingual adults to rate the similarity of pictures of objects and animals to pictures of human males or females. The target objects' labels were either grammatically masculine in Spanish and feminine in German or vice versa. All participants were tested in their L2, English, which does not mark grammatical

[1] This also occurs in English, where the generic terms dog (masculine), cow (feminine) and sheep (feminine) are used for both male and female animals, even though the language has terms for a female dog (bitch), male cow (bull) and male sheep (ram).

gender. Participants rated the items in line with their L1 grammatical markings, leading the authors to conclude that grammatical gender can “bias people’s memory for and descriptions of objects” (p.932).

However, since participants were asked to judge the similarity of unrelated items, the inclusion in the set of a picture of a man or a woman might have elicited explicit strategic reliance on the grammatical gender of object labels (Cubelli et al, 2011; Segel & Boroditsky, 2011). Other studies have had mixed results, with some studies finding grammatical gender to affect conceptual representation, so that objects seem to have more male or female characteristics (e.g. Bassetti, 2007; Yorkston & De Mello, 2005; Konishi, 1993), while others have found only modest effects emerging and only when the gender assignment is made explicit (e.g. Ramos & Roberson, 2010), or negligible effects (Vigliocco et al 2005; Bender et al., 2011; Costa et al., 2003).

Grammatical gender processing in tasks that do not draw explicit attention to gender (either grammatical or biological) has also been studied in bilinguals. Such research can be argued to be the strongest test of grammatical gender effects, since it effectively rules out other sources of variance such as cultural differences in notions of masculinity and femininity. Since ‘gender’ refers to a social construct (the roles, behaviours and attributes that a given society considers masculine or feminine) notions of gender vary substantially between different human societies, while ‘sex’ (the biological and physiological characteristics that define males and females) does not. Thus the contradictory findings from studies that have compared either monolingual speakers of two different languages or two groups of bilinguals who share a second language, but have different L1s, might relate in part to differences in cultural norms for what is considered masculine or feminine (Nicoladis & Foursha-Stevenson, 2012).

Some studies have investigated grammatical gender effects within the same population. Bordag and Pechmann (2007) found that participants named pictures in their L2 (German) faster when the object label had the same grammatical gender in their L1 (Czech) than when it had the opposite. A similar finding was reported by Morales and colleagues (2011), who asked participants first to name objects with a bare noun in their L2 and subsequently to produce the L1 article alone for the items previously named. Article naming latencies in the second task were longer for incongruent articles. The authors concluded that participants had inhibited L1 grammatical gender in the first task, and overcoming this inhibition in the second task was slow and effortful.

Findings of grammatical gender effects on response times have often been reported, but rarely processed alongside information about biological sex (though see Bender et al., 2011, for an exception). The present paper extends investigations of these phenomena to a perspective-taking task in which the voice of an instructor (male or female) is used to cue the participant as to which object to select. In this case, information about the biological sex of the instructor is apparent from the voice and thus available before the participant knows which object to select. If grammatical gender permeates mental representations of items that do not have biological sex and it is processed automatically, even when not required, it should affect reaction times even when attention is only obliquely drawn to such information.

Here, participants were asked to take one of two perspectives to select the correct object in an array, based on the 'Keysar' task (Keysar, Lin & Barr, 2003) in computerised form (Apperly et al, 2010).

Participants whose first language (L1) carried grammatical gender were compared to participants whose L1 did not and to native English speakers on a perspective-taking task in which they were required to follow instructions from either a male or female instructor, who took either the same or the opposite viewpoint as the participant.

The concept of a ‘bilingual advantage’ in certain cognitive tasks has been investigated in depth in the last 30 years (Ben-Zeev, 1977; Ianco-Worrall, 1972; Ricciardelli, 1992). It has been suggested that bilinguals possess enhanced executive control that manifests as faster and/or better performance under high cognitive load, such as in the Simon task (e.g. Martin-Rhee & Bialystok, 2008; Bialystok, 2006; Bialystok et al, 2005; Bialystok et al, 2004) reversing ambiguous figures (Bialystok & Shapero, 2005), the Dimensional Change Card Sort (DCCS) (Bialystok & Martin, 2004), and the Trailmaking task (Bialystok, 2010) (but see Hilchey & Klein, 2011, for a critical review). Researchers have often proposed that this advantage may arise because bilinguals have become expert at inhibiting one language in order to use another, or because they have developed enhanced cognitive monitoring systems, or a combination of the two. In support of the latter explanation, bilinguals often show enhanced speed and/or performance over monolinguals even when congruent and/or control trials are isolated from the experimental, incongruent ones (see Bialystok, Craik & Luk, 2012, for a review).

It was hypothesised that all participants would find taking the ‘Other’ perspective harder than taking their own (Apperly et al, 2010; Keysar, Lin & Barr, 2003; Wu & Keysar, 2007), and that this would be reflected in more distractor object choices and/or longer response latencies on ‘Other’ perspective trials compared to ‘Own’. However, if bilinguals are better able to inhibit irrelevant information, then bilingualism should modulate the difficulty of ignoring one's own perspective and/or the difficulty in switching perspectives (as opposed to maintaining one).

Gender recognition from voices is comparatively easy (Wu & Childers, 1991) and speech processing in the brain begins within 100-150ms after voice onset (Rinne et al. 1999). If the male or female voice of the instructor is processed in parallel with the grammatical gender of targets, participants whose L1 is gendered should show facilitation when both are congruent and the converse effect for incongruent items. However, if the voice of the instructor is processed independently of the grammatical gender information concerning target objects, then there should be no effect of congruency per se.

Method

Participants

16 native English-speakers, 16 participants whose L1 had no grammatical gender and 16 participants whose L1 carried grammatical gender from the University of Essex participant pool took part in return for either course credits or a small payment (see Appendix A for language data). For ease of reference, the first group is named No-L2, the second L1-NoGG, and the third L1-GG. A t-test confirmed that the two L2 groups did not differ significantly on English level, achieving the same mean score of 3.31/5 on the Oxford Quick Placement test (see Table 1), [$t_{(30)} = 0$, p=1].

(Table 1 about here)

Stimuli and Apparatus

In addition to the experimental task, all participants were tested on a range of background measures, including the short form (12 items in 15 minutes or less) of the Advanced Ravens Progressive Matrices (Arthur & Day, 1994). Participants for whom English was not their first language were administered the paper version of the Oxford Quick Placement test, which measures receptive grammar, vocabulary and collocation (Geranpayeh, 2003; Beeston, 2000). Full background measure data and test scores are given in Table 1.

The perspective-taking task was taken from a version of the 'Keysar' task (Wu & Keysar, 2007), previously adapted by Apperly and colleagues (2010), for computerised presentation. In this simplified version of the task the participant sees a 4x4 grid on the screen with 12 objects arrayed on the shelves, and two instructors, one on either side of the grid (see Figures 1a and 1b). The near instructor shared the participant's perspective of the grid, whereas the far instructor saw the grid from the opposite perspective. Four of the grid slots had occlusions at the back of the 'shelf', restricting the far instructor's view of the array, so that some objects could be seen only from the participant's perspective.

The grids measured approx. 8cm vertical x 12cm horizontal and were presented in the centre of a 15-inch laptop screen, to the left of the two instructors. The experiment was programmed using E-Prime (version 1) and run on Bootcamp on a MacBook Pro laptop computer. All experimental instructions were recorded by the same male and female voice (both British English native speakers).

Each trial began with an audio recording from either the male or female instructor, telling the participant to select a specified object within the array by mouse click. The participant was then required to click on the specified object as quickly as possible, taking into account the perspective of the instructor (i.e. what they could see) when they did so. Due to the occlusions in the grid, the same verbal command from different instructors could require a different object to be selected. When the object to be selected was the same from either instructor's perspective, that trial was a filler trial. For example, in Figure 1a, the command 'the top hammer' would require the same item to be selected from the perspective of either instructor. On critical trials, the to-be-selected object differed depending on whether the instructor who requested it shared the participant's perspective or not. For example, the command 'the bottom clock' would require a different selection from the female instructor's perspective than from the male instructor's perspective. Selecting an object from the perspective of the wrong instructor was classed as a distracter error.

A switch trial was one in which the instructor (and hence the perspective) changed from the previous trial, whereas a trial from the same instructor as the previous command was a non-switch trial. An own-perspective trial was one, which required the participant to take the near instructor's perspective, and an other-perspective trial required the participant to take the opposite perspective.

There were 8 different grid arrays, each displayed twice; once with a male as the same-perspective 'near' instructor and a female as the opposite-perspective 'far' instructor, and once with positions reversed. The same grid was never displayed twice consecutively. Each participant thus saw a total of 16 grids. Participants were cued as to which instructor they should follow by the voice (male or female) of the instructor. A complete set of grid examples is given in Appendix B.

There were a total of 128 critical trials, divided equally into male/female instructor, switch/non-switch and own/other trial types. In addition, there were 24 filler trials. There were four object types: clocks, cups, vases and hammers, and each object appeared three times in the array. Each possible combination of *switch*, *instructor gender*, *perspective* and *object* occurred four times in the experiment.

Procedure

Each participant was shown two example grids before the experiment began, once with the male as the near instructor and the female as the far instructor, and once vice-versa. Each participant then performed three practice trials before starting the task (two of which were filler trials, followed by one critical trial). If the last practice trial was not correct, corrective feedback was given and a fourth trial was performed. All participants successfully completed the practice trials.

The 152 trials were divided into blocks of 9-10 trials per grid. Instructors and grids with all test items appeared on the screen for 5000ms before the start of each block of trials for the participant to study. No instruction was given, but this gave time for the participant to establish the male and female instructors' positions in relation to the grid. Trials were subsequently presented at 3500ms intervals until the start of the next block of trials. No feedback was given, and the order of presentation of grids was pseudo-randomised.

For all participants the Keysar task was conducted first. Participants then did the short form of the Advanced Ravens Matrices set, and then the Oxford Quick Placement test (the latter when necessary). The total duration of the experiment did not exceed one hour.

Results

Background measures

Although the L1-NoGG group had a higher educational level than the other two groups, this was not reflected in fluid intelligence scores. A one-way ANOVA found no significant between-group (No-L2, L1-GG and L1-NoGG) differences on Ravens scores, [$F(2,45) = .253$, $p = .78$].

Accuracy

14% of experimental data was lost due to timeouts, responses occurring before voice onset, or failure to register mouse clicks. Selecting the distractor object on a critical trial was very rare (mean 2% for all participants) and these data were not considered further.

Response time data

RTs for incorrect responses and filler trials (16% of total responses) were removed from the analyses.

The effect of gender congruency

For the congruency analysis, only those trials with targets that carried either masculine or feminine gender in an L1-GG participant's L1 were analysed. Thus a participant whose first language used feminine grammatical gender for 'cup' was performing on a Congruent trial when 'cup' was spoken by the female instructor's voice, and Incongruent when it was spoken in a male voice. For an L1-GG participant whose L1 coded 'cup' as masculine, congruency was reversed. Comparison RTs were selected by matching participants from the other two Language groups in chronological order of participation, so that RTs for the same experimental trials could be cross-compared. Thus, the first L1-GG participant was matched to the first No-L2 and first L1-NoGG participants and if, for example, ‘cup’ was feminine in that participant’s L1, the RT data for ‘cup’ trials from the female instructor were compared to the same trials in the matched participants, creating two different baselines. Figure 2 illustrates the mean reaction times for each type of trial for each language group.

(Insert figure 2 about here)

A 3: Language (No-L2 vs. L1-GG vs. L1-NoGG) x 2: Congruency (Congruent vs. Incongruent) repeated measure *ANOVA* with repeated measures over the last factor revealed a significant effect of Language, [$F(2,45) = 3.315$, $MSE = 141131.47$, $p = .045$, $\eta_p^2 = .128$] and an effect of Congruency that approached significance, [$F(1,45) = 3.659$, $MSE = 20730.82$, $p = .062$, $\eta_p^2 = .075$], with longer latencies for congruent than incongruent trials, but no significant interaction, [$F(2,45) < 1$]. Pairwise comparisons using Tukey HSD found a slight trend towards slower performance by the group L1-NoGG compared to the No-L2 group ($p = .079$) and the group L1-GG (p = .075), with no significant difference between the latter two groups ($p = 1$). A within-participant t-test comparison of congruent vs. incongruent trials for the L1-GG group with grammatical gender alone found no difference between response times for Congruent and Incongruent trials, $t(15) = 1.044$, $p = .313$.

Perspective taking and switching

Figure 3 illustrates the mean reaction times for each type of trial for each language group.

(Insert figure 3 about here)

To examine the effect of perspective and switching, RTs were analysed in a 3: Language (No-L2 vs. L1-GG vs. L1-NoGG) x2: Perspective (own vs. other) x 2: Switch (non-switch trial vs. switch trial) mixed-design ANOVA with repeated measures over the last two factors. This revealed a main effect of Language, [$F(2,45) = 4.878$, $MSE = 464718.07$, $p = .012$, $\eta_p^2 = .178$]. There was also a main effect of Perspective, [$F(1,45) = 157.209$, $MSE = 1502327$, $p < .001$, $\eta_p^2 = .77$], due to longer latencies on other than own perspective trials. There was no significant effect of Switch, [$F(1,45) = 2.302$, $MSE = 11387.92$, p>.05], and no significant two-way interactions (Switch x Perspective, [F(1,45) ,1], Language x Perspective, [F(2,45) ,1], Language x Switch, [F(2,45) ,1], or three-way interaction, [F(2,45) ,1].

Pairwise comparisons using Tukey HSD revealed that L1-NoGG were slower than both the No-L2 ($p = .036$) and the L1-GG ($p = .019$), but the latter two groups did not differ significantly from each other ($p = .96$).

To check whether within-group variability in L2 proficiency might have masked any bilingual advantage, we compared native English speakers just to those whose L1 and L2 were most balanced from (8 from each bilingual group) in a 2: Bilingualism (lower vs. higher) x 2: Perspective (own vs. other) x 2: Switch (non-switch vs. switch) mixed-design ANOVA with repeated measures over the last two factors, which found a main effect of Perspective, [$F(1,30) = 119.709$, $MSE = 1007547.79$, $p < .001$, $\eta_p^2 = .8$], owing to longer latencies on other perspective trials.

There was no significant main effect of Switch, [$F(1,30) = 2.875$, $p = .1$], and no significant 2 or 3 way interactions [all [$F(1,30) < 1$].

Critically, there was no significant effect of bilingualism level [F(1,30) <1], nor did the two groups did not differ on their nonverbal intelligence scores, [$t(30) = -1.742$, $p = .092$]. Figure 4 illustrates the mean response times for each type of trial for these two language groups.

(Insert Figure 4 about here)

Finally, the possibility that reading direction in one's first language might have affected response times on trials on the right or left edges on the grid was explored. Mean response times for trials involving a correct response on the four leftmost and four rightmost grid squares were aggregated each participant and compared. A 3: Language (No-L2 vs. L1-GG vs. L1-NoGG) x 2: Edge (Left vs. Right) mixed design ANOVA with repeated measures over the last factor found a main effect of Edge, [$F(1,45) = 37.310$, $MSE = 185563.66$, $p = <.001$, $\eta_p^2 = .453$], due to slower responses on the right side, consistent with a left-to-right reading direction. There was also a significant main effect of Language, [$F(2,45) = 3.498$, $MSE = 169545.88$, $p = .039$, $\eta_p^2 = .135$], which appeared to mirror the response time pattern for the grid as a whole. Importantly, there was no interaction between Language and Edge, [$F(2,45) = 0.131$, $MSE = 652.28$, $p =.877$]. Pairwise comparisons using Tukey HSD found a trend towards slower performance of the group L1-NoGG compared to the No-L2 ($p = .062$) and L1-GG ($p = .074$). The groups No-L2 and L1-GG did not differ significantly ($p = 1$). This is consistent with slower responses overall by the L1-NoGG group.

Discussion

Accuracy on the task was high (84%) in all groups. All participants were faster to respond on trials that required them to take their own perspective than on trials that required them to take the perspective of an opposite instructor, but bilingualism did not appear to modulate this effect. The group of speakers of an L1 without grammatical gender performed more slowly overall than the two other groups, but this was not specific to any particular condition. Contrary to expectation, participants who's L1 carried grammatical gender neither performed more quickly than controls on congruent trials nor more slowly on incongruent trials. This suggests that there was no interaction between the biological sex of the instructor and the grammatical gender assigned to the object that participants were required to select. It may be that the grammatical gender of target objects only interacts with conflicting grammatical gender and not with the biological sex of the instructor. We take the present result as suggesting a boundary that grammatical gender effects do not cross.

All participants performed the task in English, and it remains possible that an interaction between sex of instructor and grammatical gender might be revealed if participants performed in their L1. However, the present study set out to test the effects of L1 grammatical gender on an L2 task precisely to examine how pervasive such effects might be. Implicit effects of L1 on L2 have been found previously (e.g. Bordag & Pechmann, 2007; Morales et al, 2011; Lemhofer et al., 2008). It thus seemed plausible to expect that grammatical gender effects should arise even when participants were instructed in L2 English.

Alternatively grammatical gender effects may be reduced or absent when no linguistic output is required. At no point on the perspective-taking task was a participant required to produce language, only to process it. However, studies such as those by Cubelli and colleagues (2011) with monolinguals and Bassetti (2007) with bilingual children did find effects consistent with grammatical gender in a button-pressing task, and grammatical gender has been found to influence children as young as two years old on a looking task, with no linguistic output required (Bobb & Mani, 2013). Thus the lack of linguistic output alone seems unlikely to explain the absence of an effect in the present study.

As predicted, trials from another's perspective took longer than trials from the participant's own perspective. This is consistent with the findings of a number of recent studies using a similar paradigm (Keysar et al., 2003; Wu & Keysar, 2007; Apperly et al., 2010). The lack of any interaction with Language suggests that both bilingual groups found taking another's perspective as difficult as monolinguals, at least in the current task. Nor could the lack of bilingual superiority be due to heterogeneity of L2 ability (Bialystok et al., 2004; Bialystok & Barac, 2012; Carlson & Meltzoff, 2008; Bialystok & Majumder, 1998), since even the most balanced Bilinguals performed at the same level as native English speakers.

Alternatively, the present paradigm may not have been sufficiently demanding for a bilingual advantage in executive functioning to be revealed, since it used a reduced number of items (4) and accuracy rates were high. Future research with time pressure and additional items might nevertheless reveal some differences.

In the present study bilinguals did not show superior performance on a visual perspective-taking task. At the same time there was no evidence of overlap between grammatical gender and biological sex, in a paradigm where strategic reliance on grammatical gender would be unhelpful to performance. Studies that have shown evidence for such an effect may have required more explicit use of gender information. The current finding is consistent with research by Bender and colleagues (2011), who found neither accuracy nor response latency priming effects of male and female pictograms on word/nonword judgments of inanimate object labels in German. As such, the present study may establish a limit to the relationship between grammatical gender and biological sex.

References

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A. & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Journal of Experimental Psychology*, *63(6)*, 1201-1217.

Athanasopoulos, P., Damjanovic, L., Kracjiova, A., & Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: The case of Japanese blues. Bilingualism: Language and Cognition, 14(1), 9-17.

Arthur, Winfred JR., & Day, D. V. (1994). Development of a Short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, *54(2)*, 394-403.

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

Beeston, S (2000): The UCLES EFL item banking system. *Research Notes*, *2*, 11-13

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

Ben-Zeev, Sandra (1977). The Influence of Bilingualism on Cognitive Strategy and Cognitive Development. *Child Development*, *48*, 1009-1018.

Bialystok, E. (1999). Cognitive Complexity and Attentional Control in the Bilingual Mind. *Child Development*, *70(3)*, 636-644.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60(1)*, 68-79

Bialystok, E. (2010). Global-local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology, 46*, 93–105.

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Disassociating advantages for metalinguistic awareness and executive control. *Cognition, 122(1),* 67-73.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effects of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24,* 40–49.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19,* 290 –303.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16(4)*, 240-250.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19(1),* 69-85.

Bialystok, E. & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7(3)*, 325-339.

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: the effect of bilingualism on reversing ambiguous figures. *Developmental Science, 8(6),* 595-604

Bobb, S., & Mani, N. (2013). Categorising with gender: Does implicit grammatical gender affect semantic processing in 24-month-old toddlers? *Journal of Experimental Child Psychology, 115,* 297-308

Bordag, D. & Pechmann, T. (2007).Factors influencing L2 gender processing. *Bilingualism: language and Cognition, 10(3)*, 299-314

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11(2),* 282-298.

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

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

Cubelli, R., Paolieri, D., Lotto, L., & Job, R., (2011). The effect of grammatical gender on object categorisation. *Journal of Experimental Psychology: Learning, memory and cognition*, *37(2),* 449-460.

Geranpayeh, A (2003). A quick review of the English quick placement test. *Research Notes*, *12*, 8-10.

Heider, E. R., & Olivier, D.C. (1972) The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, *3*, 337-354.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin Review*, *18*, 625-658.

Ianco-Worrall, Anita D. Bilingualism And Cognitive Development. *Child Development*, *43*, 1390-1400.

Keysar, B., Lin, S., and Barr, D.J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25-41

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

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*, 312-330.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11(1),* 81-93.

Morales, L., Paolieri, D., & Bajo, T. (2011). Grammatical gender inhibition in bilinguals. *Frontiers in Psychology*, *2*(article 284)

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

Paolieri, D., Lotto, L., Leoncini, D., Cubelli, R., & Job, R. (2011). Differential effects of grammatical gender and gender inflection in bare noun naming. *British Journal of Psychology*, *102*, 19-36.

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 20-fifth annual meeting of the Cognitive Science Society*. Boston: Cognitive Science Society.

Ramos, S. & Roberson, D. (2010). What constrains grammatical gender effects on semantic judgments? Evidence from Portuguese. *Journal of Cognitive Psychology*, *23(1),* 102-111

Ricciardelli, A. L. (1992). Bilingualism and cognitive development in relation to threshold theory. *Journal of Psycholinguistic Research, 21(4),* 301-316.

Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., Bertrand, O., and Naatanen, R. 1999. Analysis of speech sounds is left-hemisphere predominant at 100–150ms after sound onset. *NeuroReport, 10,* 1113–1117.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2004). The development of color categories in two languages: A longitudinal study.

Journal of Experimental Psychology: General, *133*, 554–571.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Colour categories in Himba: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*, 378–411.

Roberson, D., Davis, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129(3),* 369-398.

Saalbach, H., Imai, M., & Schalk, L. (2012).Grammatical gender and inferences about biological properties in German-speaking children. *Cognitive Science, 36*, 1251–1267.

Segel, E., & Boroditsky, L. (2011). Grammar in art. *Frontiers in Psychology, 1,* 244.

Sera, M., Elieff, C., Forbes, J. N., Burch, M. C., Rodrıguez, 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*, 377–397.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the investigation of language and thought*, (pp.157-191). Cambridge, MA: MIT Press

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

Whorf, B.L. (1956) The relation of habitual thought and behavior to language. In: *Language, Thought and Reality: Essays by B.L. Whorf.* J.B. Carroll (ed.) Cambridge: M.I.T. Press.

Wu, K. & Childers, D. G. (1991). Gender recognition from speech. Part 1: Coarse analysis. *Journal of the Acoustical Society of America, 90(4),* 1828-1840.

Wu, S. & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18(7)*, 600-606.

Yorkston, E., & De Mello, G. E. (2005). Linguistic gender marking and categorization. *Journal of Consumer Research, 32,* 224 – 234.

Table 1: Mean demographic data, educational level and measures of non-verbal intelligence (raw scores on short form of Ravens) and English (OQPT) for the three groups of participants tested.

	N (Females)	Mean age (SD/range)	Mean Education (1= in high school, 2= in degree, 3= in masters, 4= in PhD)	Ravens Advanced Progressive Matrices (max 12)	English level (Max 5)	L2 score Max (5)
Participants with L1 with grammatical gender (L1-GG)	16 (11)	21.31 (3.07/19-29)	2.06 (.57)	7.13 (1.4)	3.3 (1.1)	3.4 (1.1)
Participants with L1 without grammatical gender (L1-NoGG)	16 (12)	25.1 (6.4/18-42)*	2.75 (.68)	7.31 (3.7)	3.3 (1.7)	3.3 (1.7)
Native English speakers without L2 (No-L2)	16 (8)	19.8 (1.3/18-22)	2.0 (0)	6.63 (2.9)	5.0 (0)	.0.3 (.6)
Bilinguals	16 (12)	22.8 (4.5/19-34)	2.44 (.73)	8.25 (2.3)	4.44 (.629)	4.5 (.516)

Figure Captions:

Figure 1a and 1b. Examples of the sixteen stimulus grids in the set. The darker squares contained an occluding back that prevented the instructor on the opposite side of the grid from observing the contents of the shelf.

Figure 2. Mean response times for trials on which the biological sex of the instructor was either congruent or incongruent with the grammatical gender of the target object (for L1-GG participants) for native English speakers, Bilinguals whose L1 marked grammatical gender and Bilinguals whose L1 did not mark grammatical gender

Figure 3. Mean response times for trials on which participants had either to maintain or switch to their own perspective or that of the opposite viewpoint for native English speakers, Bilinguals whose L1 marked grammatical gender and Bilinguals whose L1 did not mark grammatical gender.

Figure 4. Mean response times for trials on which participants had either to maintain or switch to their own perspective or that of the opposite viewpoint for native English speakers and bilinguals.

Figures 1a and 1b

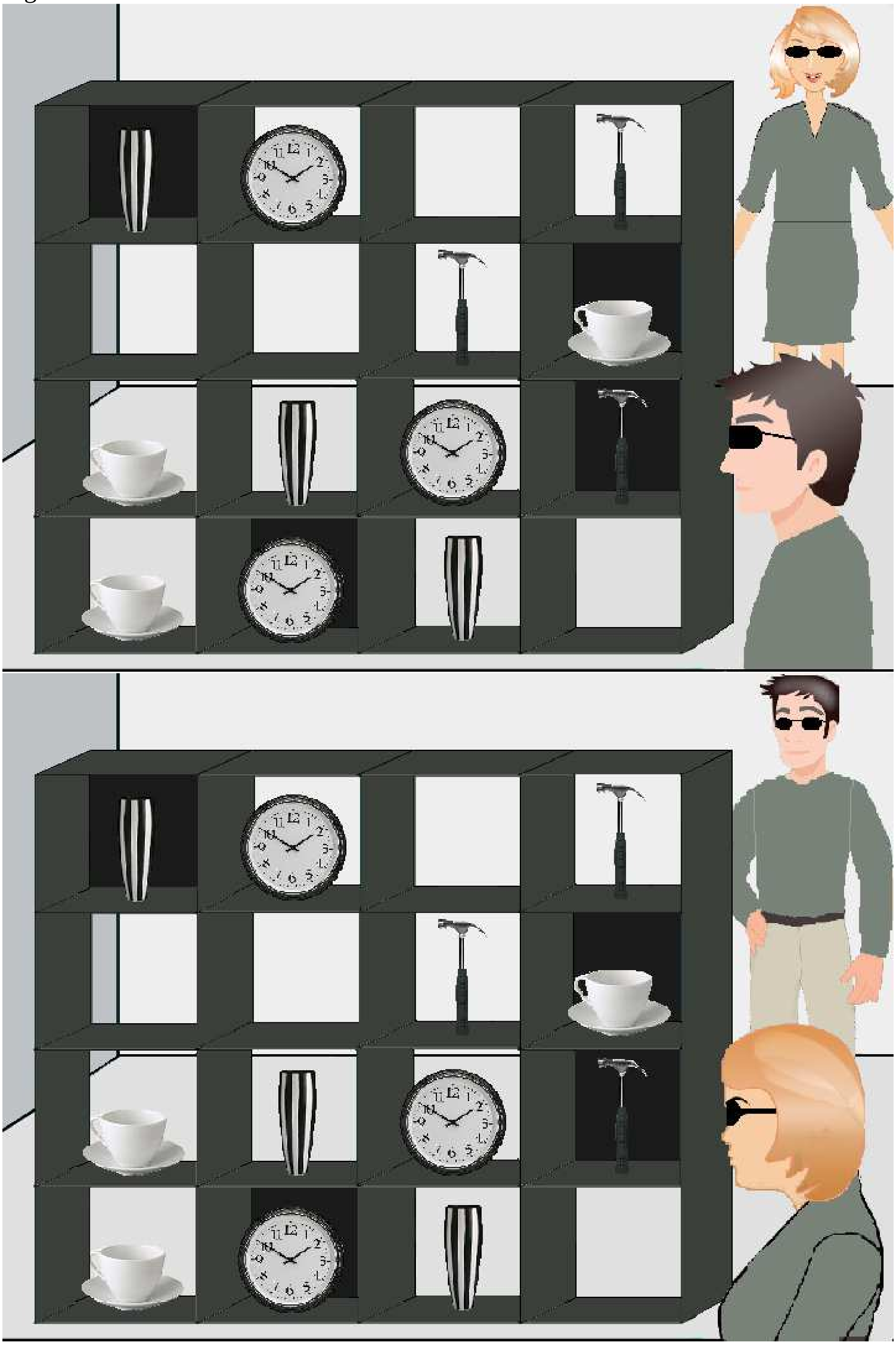


Figure 2

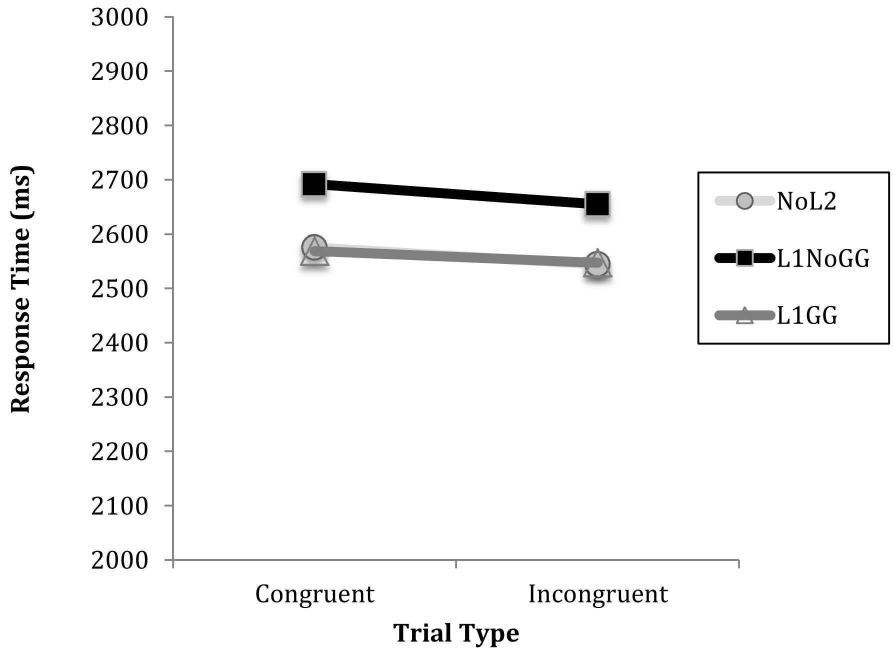

3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
Response Time (ms)
Congruent
Incongruent
Trial Type
NoL2
L1NoGG
L1GG

Figure 3

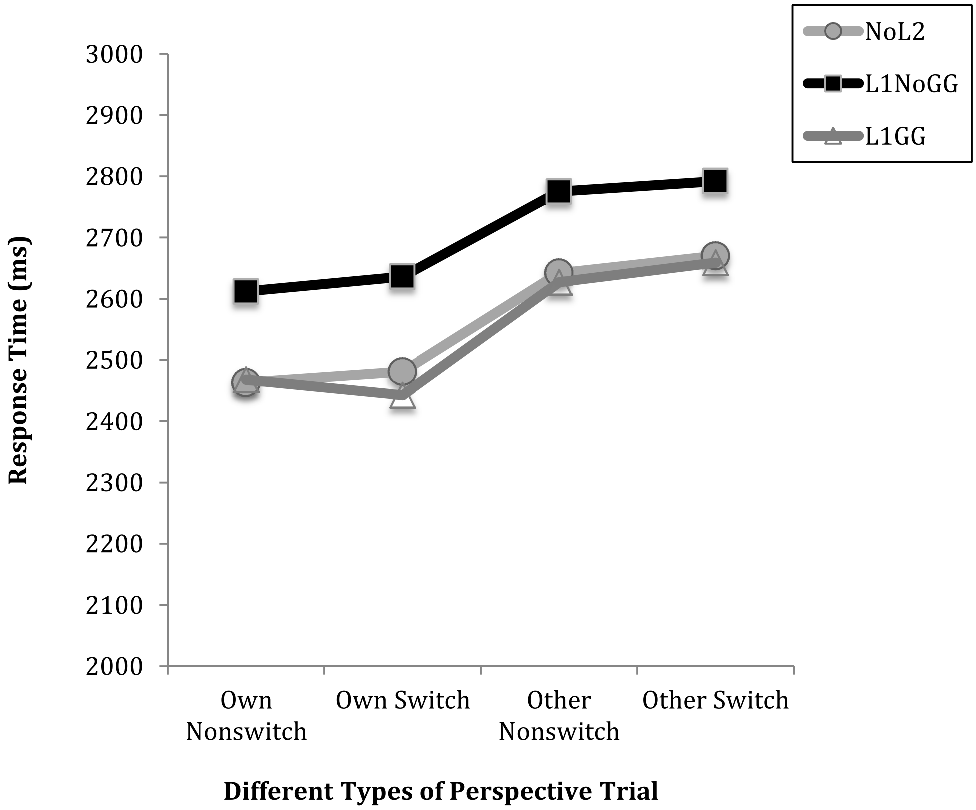
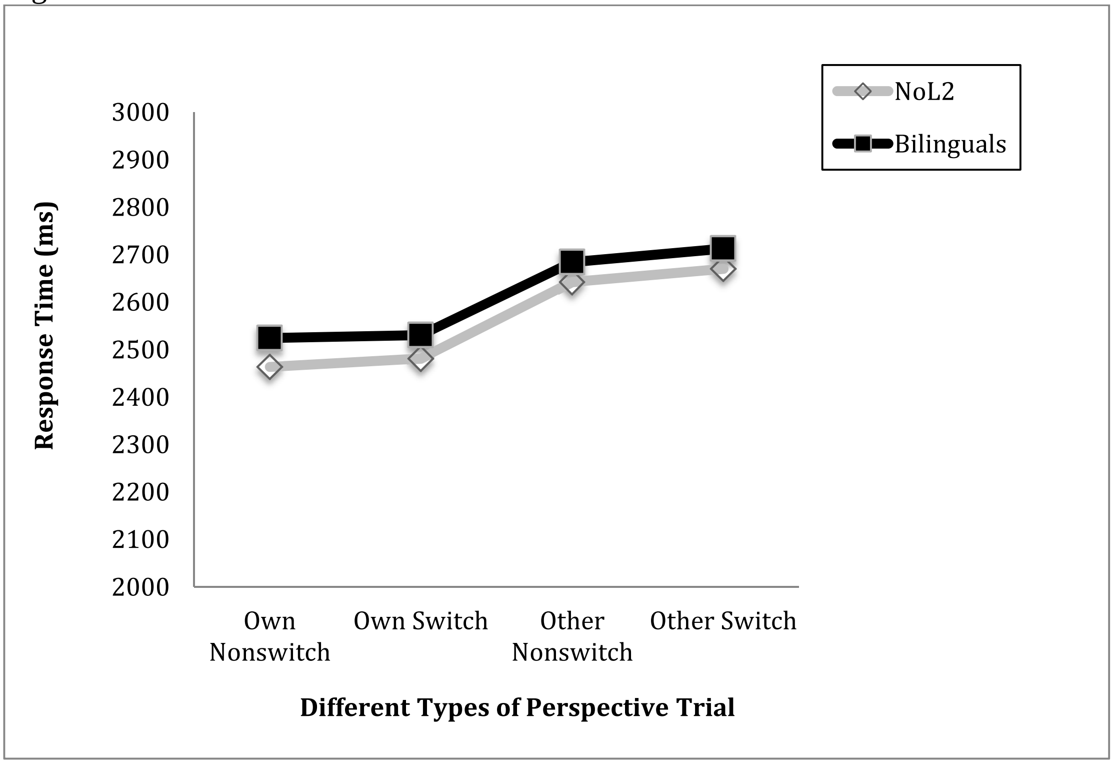

NoL2
L1NoGG
L1GG
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
Response Time (ms)
Own Nonswitch
Own Switch
Other Nonswitch
Other Switch
Different Types of Perspective Trial

Figure 4


NoL2
Bilinguals
Response Time (ms)
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
Own Nonswitch
Own Switch
Other Nonswitch
Other Switch
Different Types of Perspective Trial

Appendix A

Number of participants speaking languages with grammatical gender, and the assignment of grammatical gender to the objects in the task

Language	*N*	*Clock*	*Cup*	*Vase*	*Hammer*
Romanian	7	(neut)	fem	fem	(neut)
Lithuanian	2	masc	masc	fem	masc
Spanish	2	masc	fem	masc	masc
Italian	2	masc	fem	masc	masc
Bulgarian	1	masc	masc	fem	masc
Latvian	1	masc	fem	fem	masc
Ukrainian	1	(neut)	fem	fem	fem

Languages of participants in group L1-NoGG

Language	*N*
Chinese	5
Vietnamese	3
Korean	1
Urdu	1
Yoruba	1
Twi	1
Malay	1
Japanese	1
Swahili	1
Luganda	1

Languages of participants in group ‘Bilinguals’

Language	*N*
Romanian	4
Chinese	2
Spanish	1
Latvian	1
Spanish	1
Malay	1
Japanese	1
Vietnamese	1
Korean	1
Swahili	1
Yoruba	1
Luganda	1

Appendix B

All 16 experimental grids