

Metacognitive processing in language learning tasks is affected by bilingualism

Mikhail Ordin^{1,2}

Leona Polyanskaya¹

David Soto^{1,2}

¹BCBL – Basque Centre on Cognition, Brain and Language

²IKERBASQUE – Basque Foundation for Science

Acknowledgements:

The research was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) through the “Severo Ochoa” Programme for Centres/Units of Excellence in R&D (SEV-2015-490), and project grants RTI2018-098317-B-I00 to MO and PSI2016-76443-P to DS, by the Basque Government grant PI-2017-25 to DS. LP was supported by the European Commission with the Marie Skłodowska-Curie Research Fellowship.

Abstract

We assessed the effect of bilingualism on metacognitive processing in the artificial language learning task, in two experiments varying in the difficulty to segment the language. Following a study phase in which participants were exposed to the artificial language, segmentation performance was assessed by means of a dual forced-choice recognition test followed by confidence judgments. We used a signal detection approach to estimate type-1 performance (i.e., the participants' ability to discriminate statistical words vs. foils constructed from the same syllables), and type-2 metacognitive performance (i.e. the ability to discriminate the correctness of the type-1 decisions by confidence ratings). The results showed that bilinguals and monolinguals do not differ in type-1 recognition performance, but across the two experiments metacognitive performance was higher in bilinguals compared to monolinguals. The results show that bilingualism improves metacognitive evaluation of performance in linguistic domains. We suggest that the improvement in metacognitive performance stems from bilinguals' enhanced error monitoring abilities in language domain, which is also modulated by individual experience.

Keywords: statistical learning, speech segmentation, metacognition, bilingualism, decision confidence.

Introduction

Metacognition is the ability to monitor and evaluate one's own cognitive performance. This ability can be quantified objectively by using signal detection theoretic approaches that model the relationship between response confidence and task correctness (Maniscalco and Lau, 2014; Fleming, 2016; Fleming & Lau, 2014). Metacognition has been the subject of a much recent investigation in memory and perceptual domains (Nelson, & Narens, 1990; Song, Kanai, Fleming, Weil, Schwarzkopf, & Rees, 2011; Fleming, Ryu, Golfinos, & Blackmon, 2014; Fitzgerald, Arvaneh, Dockree, 2017). By comparison, metacognitive processes have received considerably little attention in language research. A recent study by Folke, Ouzia, Bright, De Martino and Filippi (2016) indicated a bilingual disadvantage in metacognitive sensitivity in a simple, low-level perceptual discrimination task that did not involve linguistic information. Given the lack of studies regarding the effect of bilingualism on metacognitive performance during linguistic tasks, we sought to test the hypothesis that bilingualism influences metacognitive performance in language domains, namely, in the context of an artificial language learning paradigm.

Statistical learning is the ability to extract distributional regularities from perceptual input and use them to process further input. Statistical learning is considered a domain-general mechanism operational across sensory modalities (see Armstrong, Frost & Christiansen, 2017). This mechanism is also at work for detecting statistical cues in the acoustic signal, including but not limited to computing transitional probabilities (TPs) between syllables, which is one of the pre-requisites for speech segmentation, i.e., splitting a continuous acoustic signal into sequential discrete constituents.

Previous research have not observed performance differences in statistical learning tasks between speakers of various native languages (see Saffran, Newport, & Aslin, 1996 for English; Perruchet, Poulin-Charronnat, 2012 for French, Wang & Saffran, 2014 for Mandarin, Ordin, Polyanskaya, Laka, & Nespors, 2017 for Spanish, Italian, German and Basque). Moreover, the core learning ability to extract and employ distributional cues for processing the sensory input is not affected by proficiency in multiple languages (Ordin et al., 2017; Poeppel & Weiss, 2016; Yim & Rudoy, 2013). Although bilinguals are indeed exposed to multiple sets of distributional cues and rules, this exposure is unlikely to affect the fundamental principle of statistical learning (Gallistel, Mark, King & Latham, 2001; Poeppel & Weiss, 2016). The only few studies that report the bilingual advantage in statistical learning tasks used material with multiple cues, including prosodic cues in the segmentation tasks, and prosody is indeed more likely to be processed via the phonological filter of the native language or languages (Bartolotti, Marian, Schroeder, Shook, 2011; Wang &

Saffran, 2014). Also, adding interference and thus increasing the load on selective attention allows bilinguals to employ their experience with multiple inputs having multiple cues that need to be weighted differently depending on the current state of the ever-changing environment (Bartolotti et al., 2011; Kovac & Mehler, 2009; Wang & Saffran, 2014).

Importantly, a number of studies have demonstrated a correlation between individual performance on statistical learning tasks and the comprehension of natural language sequences (Misyak & Christiansen, 2012; Conway, Bauerschmidt, Huang, & Pisoni, 2010). Individual differences in statistical learning can be linked to the development of language processing skills, linguistic proficiency, language impairments and reading difficulties associated with lower control over executive functions and phonological working memory (Conway & Pisoni, 2008; Evans, Saffran, Robe-Torres, 2009; Majerus, Poncelet, van der Linden, & Weekes, 2008; Vicari, Marotta, Menghini, Molinari & Petrosini, 2003). Nation and McLaughlin (1986) showed that multilinguals might outperform bilinguals and monolinguals in statistical learning tasks (Markovian artificial grammar learning) if the learning task was implicit (i.e., participants had to learn the grammar from exposure, but were not informed that the input was constructed according to the rules). In explicit learning of the same material, i.e., when the participants were informed upfront that they would be shown artificial words derived according to statistical rules and were asked to extract the rules, multilingual advantage was not observed.

These results support the view that statistical learning is not a single, fixed mechanism, but rather a set of cognitive processes inherent to attention and memory (Thiessen, 2017) that are modulated by individual experience and by one's own metacognitive processing skills that monitor learning and the outputs of the learning process. Simultaneous acquisition of multiple linguistic systems hones many cognitive processes related to phonological memory (Majerus et al., 2008) and attention (Hernández et al., 2011), which may affect the conscious awareness of what is being learnt and the ability to evaluate the sense of cognitive performance in the test. This study is devised to test this hypothesis by using an artificial language learning paradigm.

The present study has theoretical implications for understanding both metacognition and bilingualism. The project is relevant to identify the factors that influence metacognitive confidence during task performance. Understanding the role of linguistic factors (e.g. bilingualism) on metacognition would also provide insight into the extent metacognition is influenced by experience dependent changes. Recent research indicates that metacognitive skills are subject to training effects in a specific task and more critically that metacognitive training effects can

generalise across cognitive domains (Carpenter et al., 2019). Regarding bilingualism, the project is relevant to understand whether and how bilingualisms influences cognition. Whether or not bilingualism is associated with benefits in cognitive performance even in non-linguistic tasks such as executive control has been the subject of much debate (Duñabeita & Carreiras, 2015; Paap Johnson & Sawi, 2015). Our study however deals with the role of bilingualisms in a language context. While no previous research revealed differences between monolinguals and bilinguals in first-order segmentation performance in the language learning paradigm, we here examine for the first time differences in second-order metacognitive performance between bilinguals and monolinguals.

Experiment 1

Methods

Participants

48 Spanish monolinguals from Murcia and 48 Basque-Spanish bilinguals from San Sebastian were recruited. The bilinguals started acquiring Basque and Spanish before the age of two, reported using the standard Batua dialect, functioning daily in both languages and being immersed in bilingual environment, and were equally proficient in both languages, as measured by the BEST test (de Bruin, Carreiras, & Duñabeitia, 2017). The participants from both groups were matched in the level of education and in age (20-33 years old). Previous studies have shown that the socio-economic background of Spanish students who have chosen to go to university is similar nationwide (Lopez, 2009), and the curricula and the admission criteria for public universities in the Basque Country and in Murcia are also similar (CRUE, 2018). Thus, it is reasonable to assume that the samples of bilinguals and monolinguals do not significantly differ in terms of socio-economical and educational variables.

Materials

We used 5 vowels (/a, o, u, e, i/) and 6 consonants (/k, t, p, k, l, m, r(trill)/). These phonemes were selected because they exist in both Basque and Spanish – native languages of the experimental participants. We concatenated these sounds to construct two sets with six tri-syllabic nonsense words each. Further on, we will refer to these sets as Language One (L1) and Language Two (L2). L1 includes the words *kitepo, maleru, purime, ketapi, raloku, moluto*. L2 includes the words *tokape, romuli, tamire, pituko, malupa, lariku*. Several studies have demonstrated that lexical processing is supported by consonants, and people probably rely on TPs between adjacent consonants more than on the TPs between vowels or syllables. Therefore, the words for our experiment were constructed so that the 3-consonant combinations for each word were unique. Each consonant was used three times within each artificial

language and was not repeated within one word. The syllables within each artificial language were unique. Also, we tried to maximize and randomize the use of vowels within and between languages.

We used MBROLA with diphone databases ES1 and ES2 for stimuli preparation. Each artificial language was created by concatenating the statistical words so that each word is repeated an equal number of times (N=200) and the same word is not repeated successively. In each artificial language, the TPs between syllables within words were 1.0 and between syllables straddling the word boundaries were 0.16. No pauses were introduced within the acoustic stream. Each artificial language was synthesized with the ES1 and ES2 voice. Each consonant was set to 100ms and each vowel was set to 140ms in duration. The pitch contour was flat and set to 100Hz for voice ES1 and 200Hz for voice ES2. The beginning and the end of each stream was ramped in intensity for the initial and final 5 seconds, to prevent the listeners from clearly identifying the initial and the final syllables. Intensity level was normalized at 70dB. Audacity was used to create stereo WAV files with different artificial languages presented into each ear. The onsets of the acoustic streams were synchronized. Four different files for dichotic presentation of artificial languages were made, counterbalancing the language (L1 and L2) * voice (ES1 and ES2) * channel (Right and Left).

The stimuli were presented in PsychoPy(v.1.80.04) via Etymotic MC5 headphones, using an identical hardware and software setup in both experimental locations.

Experimental Procedure

Two artificial languages were presented concurrently using dichotic presentation of each language to each ear. The experiment comprised separate study and test phases. In the study phase, the listeners were told they were going to listen to two creatures speaking simultaneously in the same extraterrestrial language (one per ear), but using different words. The task was to detect and memorize as many words of the extraterrestrial language as possible. Immediately after familiarization, the test phase assessed the segmentation performance by means of a dual forced-choice recognition task, in which statistical words were paired with random combinations of syllables of the artificial language. The items within pairs were separated by 500ms. The test pairs were presented either to the left or to the right ear. Each word was used in 4 pairs pitted against different random syllabic concatenations, counterbalancing the position of the item in pair and the ear (72 pairs in total). Presentation of each pair was followed by two questions. Participants were asked to choose whether the first or the second item in each pair was the word from the extraterrestrial language, and then – to probe differences in metacognitive processes between monolinguals and bilinguals – to indicate how confident they were in their answer on a 4-point scale: 1-guess, 2-little confident, 3-fairly

confident, 4- sure. We encouraged the participants to use the whole range of confidence ratings. The study (Experiments 1 and 2) was approved by the ethical committee of the BCBL (Ethics Approval Number in the institutional records: 18042018M).

Signal Detection Theory (SDT) approach

Metacognition is typically modeled as a two-stage process involving a type-1 decision regarding a target item, followed by a type-2 decision involving a metacognitive judgment of one's confidence in the type-1 decision. Recent approaches within signal detection theory allow the effective isolation of metacognitive ability (the relationship between type-1 performance and confidence) from fluctuations in type-1 performance (Maniscalco & Lau, 2012; 2014; Fleming & Lau, 2014). This approach basically involves estimating sensitivity to the type-1 information available to make type-2 decisions by maximizing the likelihood of the observed type-2 ROC for each participant. This requires parameter estimation of type-2 sensitivity predicted from type-1 model by fitting type-1 model to the observed type-2 data using log-maximum-likelihood estimation method (Maniscalco & Lau, 2014). This sensitivity measure, referred to as meta- d' , is in the same units as the measure estimating the first-order sensitivity, or d' . Therefore, a straightforward way to calculate the metacognitive efficiency is the M-ratios (meta- d'/d'). M-ratios show individual metacognitive performance controlling for the individual level in type-1 performance. Here we compared metacognitive performance between monolinguals and bilinguals using the hierarchical Bayesian estimation method (Fleming & Daw, 2017), which improves estimation when the amount of available trials is limited (Fleming, 2017).

To use this approach, type-1 hits were defined as the trials in which the target word in the test pair appeared in the first position and the participant responded 'first', while false alarms were defined as the trials in which the target appeared in the second position within the test pair and the participant responded 'first'. Correct rejections were defined when as the trials in which the target word in the test pair appeared in the second position and the participants responded 'second' (i.e., correctly rejecting the option that the signal – target word – was on the first position), and misses were defined as the trials in which the target appeared in the first position, and the participant responded 'second' (i.e., missing to spot the target in the first position). Input vectors containing the total number of responses as a function of confidence were derived from each participant. These vectors were used to estimate meta- d' and M-ratios (meta- d'/d') using the code available on (<https://github.com/metacoglab/Hmeta-d>, Fleming, 2017).

Results and Discussion

To estimate the segmentation performance (i.e., first-order performance), we compared the accuracy of segmentation of the first and second artificial languages, artificial languages presented in ES1 and ES2 voices, and artificial languages presented in the right and in the left ear. To test the overall success of the segmentation, we carried out a one-sample t-test for the proportion of correct responses against the 50% chance level, and compared the proportion of correct responses between bilinguals and monolinguals to test whether one group outperforms the other (**Figure 1A**). One-sample t-test showed that the proportion of correct responses was significantly above chance both in the group of monolinguals, $t(47)=6.323$, $p<.0005$, Cohen's $d=.91$ and bilinguals, $t(47)=7.161$, $p<.0005$, Cohen's $d=1.05$. The proportion of correct responses is not different between bilinguals and monolinguals, $t(94)=1.287$, $p=.201$, Cohen's $d=.26$. All reported t-tests are 2-tailed.

Type-1 sensitivity did not differ between the groups of bilinguals and monolinguals, $t(94)=1.306$, $p=.195$, Cohen's $d=.27$ and was significantly above zero, $p<.0005$ for both groups, which again confirms that bilinguals and monolinguals do not differ in the effectiveness of fundamental SL mechanism. We then assessed metacognitive efficiency by estimating M-ratios (**Figure 1B**). Metacognitive efficiency is significantly higher in the group of bilinguals compared to monolinguals, $t(94)=6.166$, $p<.0005$, Cohen's $d=1.26$, revealing a bilingual advantage in metacognitive efficiency.

Finally, we tested whether bilinguals and monolinguals differ in metacognitive bias, i.e., the tendency to assign higher or lower confidence ratings overall. We averaged the confidence ratings across trials for each participant (**Figure 1C**), and compared the overall confidence between bilinguals and monolinguals. The test revealed no difference in metacognitive bias between groups, $t(94)=1.09$, $p=.279$, Cohen's $d=.22$ (after correction of degrees of freedom to account for a slight inequality of variance, as revealed by the Levene's test, $F=6.1$, $p=.015$, the result pattern remains unaffected, $t(85.27)=1.09$, $p=.279$, Cohen's $d=.22$). This result shows that the effect of bilingualism in metacognitive efficiency is not explained by the tendency of assigning higher or lower confidence ratings by participants from different groups.

Taken together, these results indicate that segmentation of dichotically presented artificial languages was successful across the groups but similar between bilinguals and monolinguals. Accuracy in type-1 performance and metacognitive bias – mean confidence – were matched between the groups. Metacognitive processing by bilinguals and monolinguals was compared using the SDT framework, which allows estimating metacognitive efficiency

unbiased by the *individual* level of type-1 performance (Fleming & Daw, 2017). We observed that bilinguals exhibited an advantage in metacognitive efficiency in segmentation of a continuous artificial language in auditory modality into discrete sequential constituents.

Experiment 2

In the first experiment, the dichotic presentation imposed a great load in low-level auditory processing. In the current Experiment 2, we wanted to replicate the results of the first experiment in the absence of the perceptual load caused by dichotic stimuli presentation.

Method

Participants

We recruited 36 Spanish monolinguals from Murcia and 36 Basque-Spanish bilinguals from San Sebastian with the same sociolinguistic profile as in experiment 1. Here we set 36 participants as the target sample because the amount of conditions that had to be counterbalanced was less than in Experiment 1.

Materials

We used the same statistical words as in experiment 1, but all the words (N=12) were concatenated into the same continuous stream of syllables presented simultaneously into both ears. Each word is repeated an equal number of times and the same word is not repeated successively. Dichotic presentation in experiment 1 was implemented to increase the complexity of the task. In experiment 2, we decided to increase the complexity by diminishing the number each was repeated in the stream (N=125). The stimuli for prepared with MBROLA (ES2 voice), the statistical regularities, durational and F0 parameters and in the familiarization stream were equal in both experiments. The experimental hard- and software setup in experiment 1 and 2 was the same.

Experimental Procedure

The procedure in experiment 2 (familiarization phase followed by recognition test) was the same as in experiment 1. SPSS was used to perform the statistical analyses in both experiments.

Results and Discussion

One-sample t-tests showed that the proportion of correct responses was significantly above chance both in the group of bilinguals, $t(35)=10.193$, $p<.0005$, Cohen's $d=1.7$, and in the group of monolinguals, $t(35)=11.883$, $p<.0005$, Cohen's $d=1.98$. The proportion of correct responses (**Figure 2A**) is not different between bilinguals and monolinguals, $t(70)=-1.324$, $p=.19$, Cohen's $d=.31$.

To estimate metacognitive efficiency, we again first calculated type-1 sensitivity, which was not different between groups, $t(70)=.16337$, $p=.186$, Cohen's $d=.04$, and was significantly above zero for both groups at $p<.0005$. We again observed a significant difference between bilinguals and monolinguals in metacognitive efficiency, M-ratios, $t(70)=3.288$, $p=.002$, Cohen's $d=.77$, reflecting a bilingual advantage in metacognitive processing. As it is evident on **Figure 2B**, the variance in M-ratios in the group of bilinguals is substantially and statistically larger than in the group of monolinguals, $F=15.07$, $p<.0005$ (Levene's test for the equality of variance, also evident on **Figure 2B**). However, additional analyses showed there were no outliers in the distribution of the bilingual data (outliers are defined as z-transformed data points exceeding ± 3.29 , we actually did not observe z-transformed data points exceeding ± 2.58 , although in normally distributed data it is possible to have 1% of such data points during random sampling from populations). Moreover, correcting the degrees of freedom to account for the inequality of variance does not affect the result pattern, $t(37)=3.288$, $p=.002$, Cohen's $d=.77$, bilinguals exhibiting more efficient metacognitive processing than monolinguals. The result remains robust in the presence of variability difference between the groups in M-ratio scores.

Finally, we compared overall confidence by bilinguals and monolinguals (**Figure 2C**). The results, $t(70)=.595$, $p=.554$, Cohen's $d=.14$ did not show that confidence was overall higher or lower in any of the two groups.

These results replicate the pattern obtained in the first experiment, showing that the form of habituation, i.e., dichotic presentation of two artificial languages, which supposedly increases the cognitive and attentional load, was not a requirement for bilingual advantage in metacognitive processing to emerge. In order to better estimate a true effect of bilingualism on metacognitive processing in language-related tasks, we performed a meta-analysis across both studies, pooling together effect sizes (Cohen's d) from experiments 1 and 2 (**Figure 3**). The analysis was performed on unbiased values of Cohen's d , model with random effects was used because the variability in effect sizes could potentially be not only due to sampling error, but also due to other unaccounted factors, e.g., the difference in the task difficulty in experiments 1 and 2. In experiment 1, the effect size $d=1.26$, $d_{\text{unbiased}}=1.25$, $N1=N2=48$, 95% CI on d_{unbiased} for the difference in M-ratios between groups is [0.819:1.696], weight (contribution of a particular study into estimation of the true effect size) is 51.9%. In experiment 2, the effect size $d=.77$, $d_{\text{unbiased}}=.76$, $N1=N236$, 95% CI on d_{unbiased} for the difference in M-ratios between groups is [0.288:1.247], weight is 48.1%. The pooled $d_{\text{unbiased}}=1.015$, 95% CI is [0.537:1.493], suggesting that the expected difference in metacognitive efficiency

between monolinguals and bilinguals on language-related tasks is around 1 standard deviation, which is a substantial effect, consistent across studies.

Finally we also conducted Bayesian t-tests to assess the amount of evidence in favour the bilingual advantage that we report in metacognitive processing. For experiment 1, $BF_{10}=665944$ provides a decisive evidence hypothesis that the metacognitive efficiency, measured as M-ratios, is indeed different between the group of bilinguals and monolinguals. **Figure 4A** depicts the 95% credible interval for the difference: [0.774, 1.653] with M (median)=1.208. For experiment 2, $BF_{10}=20.899$ suggesting that the alternative hypothesis that bilingual metacognitive efficiency was different from monolinguals was 21 times more likely than the null, 95% credible interval for the difference: [0.227, 1.184] with M (median)=.693 (**Figure 4B**). Therefore, given the data from both experiments, there is strong evidence (Jarosz & Wiley, 2014) for bilingual advantage in metacognitive processing.

Understanding the source of the effect of bilingualism on metacognition

Here we aimed to narrow down the mechanism underlying the improvement in metacognitive processing in bilinguals. We hypothesised that bilingualism effects a change in systems for error monitoring in language tasks. Recall that successful metacognitive processing can be reflected either in the assignment of high confidence following a correct response but also to low confidence to incorrect responses (Fleming & Lau, 2014; Schwiedrzik, Singer & Melloni, 2011). Accordingly we decided to assess whether and how confidence scores varied on correct and incorrect responses separately. An effect of bilingualism in metacognitive confidence for incorrect responses would be in keeping with a role in error monitoring mechanisms. In order to test this we lumped together the data across the two Experiments and performed a repeated-measures ANOVA with *correctness* (correct vs. incorrect) as a within-subject factor and *group* (bilingual vs. monolingual) and *experiment* (exp. 1 vs. exp. 2) as between-subject factors.

The result showed that, as expected, correct responses were generally assigned higher confidence than incorrect responses, $F(1,168)=172.97$, $p<.0005$, $\eta_p^2=.507$. We also found a trend towards overall confidence being slightly higher in monolinguals than bilinguals, however the effect size was weak and on the verge of significance, $F(1,168)=4.416$, $p=.047$, $\eta_p^2=.026$. Importantly, we found a significant interaction between *group* and *correctness*, $F(1,168)=6.26$, $p=.013$, $\eta_p^2=.046$. There was no main effect of *experiment*, $F(1,168)=1.535$, $p=.217$, $\eta_p^2=.009$, no interaction between *experiment* and *group*, $F(1,168)=.019$, $p=.89$, $\eta_p^2<.0005$, and no significant three-way interaction between *experiment*, *group*, and *correctness*, showing that confidence ratings did not differ between two

experiments neither for bilinguals nor for monolinguals. Further analyses showed that the level of confidence assigned to correct responses is not different between monolinguals and bilinguals, $t(171)=-.788$, $p=.432$, Cohen's $d=.116$, unpaired t-test, two-tailed. The ratings assigned to incorrect responses by bilinguals are significantly and substantially lower than those by monolinguals, $t(171)=-2.97$, $p=.003$, Cohen's $d=.453$, unpaired t-test, two-tailed (**Figure 5A**), this pattern is consistent across both experiments. To confirm that the difference between confidence ratings assigned to correct and incorrect responses is indeed greater in bilingual population, we performed Bayesian t-test to figure out the evidence in favour of the hypothesis that this difference is greater in bilinguals than in monolinguals. $BF_{10}=9.028$ provide substantial evidence (Jarosz & Wiley, 2014) that the difference in confidence ratings assigned to correct and incorrect trials is larger in bilingual population (**Figure 5b**). Finally, we shuffled these differences in confidence ratings on correct and incorrect trials between monolinguals and bilinguals and performed an exact permutation test. The results showed that it is rather unlikely that the observed differences between groups happened by chance (exact $p=.025$). Exact permutation tests for confidence ratings assigned by bilinguals and monolinguals to correct responses showed that any observed difference between groups is very likely to happen by chance (exact $p=.571$), and the differences in the confidence ratings assigned to incorrect responses are unlikely to happen by chance (exact $p=.015$). These results show that the bilingual advantage in metacognitive processing in the language domain may stem from enhanced error monitoring processes.

We note a higher degree of variability in metacognitive efficiency in the group of bilinguals compared to monolinguals (see Figure 2B). Given that metacognitive processing may be affected by individual experience in a particular task (e.f., Rademaker & Pearson, 2012; Carpenter et al., 2019), it is possible that greater variability in the linguistic experience of the bilingual group may have resulted in greater variance in metacognitive efficiency estimates. Bilingualism is not a monolith phenomenon but rather multi-faceted. Accordingly, the Adaptive Control Hypothesis (Green & Abutalebi, 2013) proposes that bilingual individuals have different cognitive profiles relative to monolinguals based on the distinct language control processes that are engaged in dual-language contexts. Hence, bilinguals' sensitivity to errors in a language task may be modulated by the amount of exposure to dual-language contexts within the bilingual population, as opposed to a single-language context, when two languages in the bilinguals' inventory are not mixed within the same interactional situation. This would lead to inter-individual variation in the expression of metacognitive confidence following error trials. By comparison, according to the

adaptive control hypothesis (Green & Abutalebi, 2013) the group of monolinguals is more homogeneous because, by default, they are not functioning in dual-language use context.

General Discussion

We found that type-1 segmentation performance in the artificial language learning paradigm is not different between bilinguals and monolinguals. This finding is consistent across both experiments, although the dichotic listening condition was more challenging and cognitively demanding for the participants, which is evident in higher percentage of correct responses in the second experiment (around 70%) compared to that in the first experiment (around 60%). This finding confirms that the fundamental SL mechanism, namely, tracking transitional probabilities between adjacent elements for the purposes of segmenting continuous auditory input into discrete recurrent constituents, is not influenced by bilingualism, although learners and users of two or more languages need to maintain different sets of statistical probabilities present in their ambient languages. In this regard, our data is in line with earlier studies (see, for example, Ordin et al., 2017; Poepsel & Weiss, 2016; Yim & Rudoy, 2013). However, metacognitive processing is under the influence of bilingualism: metacognitive efficiency is enhanced by exposure to and use of several languages. The key finding was that the bilinguals outperform monolinguals in metacognitive efficiency. This result does not align with the study of Folke et al. (2016) in which a bilingual dis-advantage in metacognitive sensitivity was reported. These seemingly different patterns of results between our study and Folke et al. may be due to the different paradigms used. Here we used a statistical learning task related to speech segmentation while Folke's et al. (2016) used a low-level perceptual discrimination task of non-linguistic nature. We propose that there can be improvements in domain-specific (i.e. language) metacognitive processing that are driven by experience in this particular domain.

Notably, our data indicates that bilinguals assign lower confidence ratings to incorrect trials than monolinguals, while ratings on correct trials do not differ significantly between the groups. This result suggests that the differences in metacognitive processing in bilinguals vs monolinguals relates to the monitoring of error responses during the recognition of the learned artificial words. This finding helps to narrow down the mechanism underlying the improvement in metacognition in bilinguals. It is in line with previous studies of the error-related negativity (ERN), which is thought to represent an electrophysiological index of *conscious* performance monitoring. Studies of the ERN provide indirect evidence to support the general idea that bilinguals and monolinguals may differ in how they process error responses (Morales, Yudes, Gomez-Ariza, & Bajo, 2015; Kousaie & Philips, 2012). A larger

negative deflection distributed in the medial-frontal scalp area reflects a higher degree of conflict between two simultaneously activated response patterns. Smaller ERN amplitude corresponds to lower conflict condition due to higher awareness of the error being committed, while larger ERN is associated with lower awareness that the error has been committed, which leads to activation of different response tendencies (Yeung, Botvinick & Cohen, 2004). Smaller ERN, which is typically associated with better and more efficient error-monitoring system, has been detected in bilingual compared to monolingual populations (Morales et al., 2015; Kousaie & Philips, 2012).

Whether metacognition is mediated by domain-general or domain-specific mechanisms has been highly debated (Fleming et al., 2014; Fitzgerald et al., 2017; McCurdy, Maniscalco, Metcalfe, Liu, de Lange, & Lau, 2013; Miele, Wager, Mitchell, & Metcalfe, 2011; Veenman, Elshout, & Meijer, 1997; Song et al., 2011) with recent evidence pointing out for the existence of a domain general system (Carpenter, Sherman, Kievit, et al., 2019). However, it is possible that some aspects of metacognitive processing may be modulated by individual's experience in a specific domain (i.e., Rademaker & Pearson, 2012) and accordingly, we propose that the detected differences in metacognitive efficiency may relate to modulatory effects of the individual linguistic experience on the monitoring of how well statistics are recognized.

The present study also raises additional questions for future research to address whether the role of bilingualism in metacognitive processing is transferrable to other cognitive domains and/or sensory modalities. Although there is no evidence so far that individual experience in language processing can lead to enhanced metacognitive processing beyond the linguistic domain, research suggest that metacognitive skills are transferrable in healthy individuals across different tasks in different cognitive domains (Veenman, Elshout & Meijer, 1997; Song et al., 2011; Carpenter et al., 2019). Therefore, it is possible that enhanced metacognitive processing in the linguistic domain can potentially be transferred to non-linguistic tasks because the brain bases of metacognitive processing are often reported to be task-independent and operate across modalities (visual and auditory) and cognitive domains (e.g., linguistic, logical, creative artistic and others, McCurdy, Maniscalco, Metcalfe, et al., 2013).

The present research may also have societal ramifications, particularly in the context of group-based decision making. Disagreements can occur between individuals and in this case, more importance is assigned to the opinion of those who express themselves with higher confidence (Zarnoth & Sniezek, 1997). The expression of confidence at the same time can vary with factors that are not directly relevant to the decision at hand, including social, cultural and psychological (Ais, Zylberberg, Barttfeld, Sigman, 2016; Mann, Radford, Burnett, et al., 1998).

Given that multilingualism is an integral characteristic of contemporary society, it will be relevant to address how collective decisions are influenced by bilingualism, for instance, whether decisions are discussed in the context of L1 vs L2 and the ratio of bilinguals and monolinguals within the group.

Finally, the current study lays the foundation for additional research to increase our understanding of the relations between bilingualism and metacognitive processing across different task contexts, which has so far received little attention. Studying the impact of bilingualism on cognition is hampered by the need to control a number of factors that include socio-cultural variables such as the exposure to a bilingual environment. An interesting direction for future research will be to explore the role of inter-individual differences in language cognitive profiles in metacognitive processing during statistical learning tasks, namely, the percentage of use of two languages on a daily basis, whether the languages in the bilingual's inventory are separated according to the social contexts, or whether the languages are used across all contexts interchangeably. Although we tried to recruit balanced bilinguals, it may be possible that some bilinguals were more proficient in Basque than others, and the issue of relations between linguistic proficiency and metacognitive efficiency in linguistic domains also needs empirical investigations. Additional studies in this vein are also needed, for instance, to probe statistical learning mechanisms among groups of bilingual individuals that vary in the age-of-acquisition of the languages and the level of second-language proficiency.

References

- Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* 146, 377-86.
- Armstrong, B., Frost, R., & Christiansen, M. (2017). The Long Road of Statistical Learning Research: Past, Present, and Future. *Philosophical Transactions of the Royal Society of London: Biological Sciences* 372(1711).
- Bartolotti, J., Marian, V., Schroeder, S., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology* 2: 324.
- Carpenter, J., Sherman, M.T., Kievit, R.A., Seth, A.K., Lau, H. & Fleming, S.M. (2019) Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General* 148(1), 51-64.
- Conway, C. & Christiansen, M. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition* 31, 24-39.

- Conway, C. & Christiansen, M. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus specific representations. *Psychological Science* 17, 905-912.
- Conway, C., & Pisoni, D. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences* 1145, 113–131.
- Conway, C., Bauernschmidt, A., Huang, S., & Pisoni, D. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition* 114, 356–371.
- CRUE (2018). La Universidad Española en Cifras 2016-2017. Retrived on 2.05.2019 from <http://www.crue.org/SitePages/La-Universidad-Espa%C3%B1ola-en-Cifras.aspx>
- De Bruin, A., Carreiras, M., & Dunabeitia, J. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology* 8:522.
- Duñabeitia, J. A., & Carreiras, M. (2015). The bilingual advantage: Acta est fabula? *Cortex* 73, 371–372.
- Evans, J., Saffran, J., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 52, 321–335.
- Fitzgerald, L.M., Arvaneh, M., & Dockree, P.M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Conscious and Cognition* 49, 264–277.
- Fleming, S. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness* 1, 1-14.
- Fleming, S., & Daw, N. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review* 124(1), 91-114.
- Fleming, S., & Lau, H. (2014). How to measure metacognition. *Frontiers in Human Neuroscience* 8:443.
- Fleming, S.M., Ryu, J., Golfinos, J.G., & Blackmon, K.E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* 137, 2811–2822.
- Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in metacognitive processing. *Cognition* 150, 119-132.
- Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes* 27(4), 354-372.
- Green, D., & Abutalebi, J. (2013). Language control in bilinguals: The predictive control hypothesis. *Journal of Cognitive Psychology* 25(5), 515-530.
- Hernández, M., Costa, A., Fuentes, L., Vivas, L. J., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition* 13, 315–325.

- Jarosz, A.F., Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *Journal of Problem Solving* 7, 2-9.
- Kousaie, Sh., & Phillips, N. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research* 1446, 71-90.
- Kovács, A., & Mehler, J., (2009). Cognitive gains in 7-month-old bilingual infants. *PNAS* 106, 6556–6560.
- Lopez, M.R. (2009). Equality of opportunities in Spanish higher education. *Higher Education* 58, 285-306.
- Majerus, S., Poncelet, M., van der Linden, M., & Weekes, B. (2008). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition* 107, 395–419.
- Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: meta-d', response-specific meta-d', and the unequal variance SDT mode. In S. M. Fleming & C. D. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp.25-66). Springer.
- Mann, L., Radford, M., Burnett, P., Ford, S., Bond, M., Leung, K., Nakamura, H., Vaughan, G., Yang, K.-S. (1998). Cross-cultural differences in self-reported decision-making style and confidence. *International Journal of Psychology* 33(5), 325-335.
- McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K.Y., de Lange, F.P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience* 33, 1897–1906.
- Miele, D.B., Wager, T.D., Mitchell, J.P., & Metcalfe, J. (2011). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience* 23, 3620–3636.
- Misyak, J.B. & Christiansen, M.H. (2012). Statistical learning and language: An individual differences study. *Language Learning* 62, 302-331.
- Morales, J., Yudes, C., Gomez-Ariza, C., & Bajo, T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERP. *Neuropsychologia* 66, 157-169.
- Nelson, T. O. & Narens, L. (1990) Metamemory: a theoretical framework and new findings. *Psychology, Learning and Motivation: Advanced Research and Theory* 26, 125–173.
- Ordin, M., Polyanskaya, L., Laka, I., & Nespors, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory and Cognition* 45(5), 863-876.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* 69, 265–278.
- Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language* 66, 807-818.
- Poepsel, T., & Weiss, D. (2016). The influence of bilingualism on statistical word learning. *Cognition* 152, 9-19.

- Rademaker RL, Pearson J. (2012) Training visual imagery: Improvements of metacognition, but not imagery strength. *Frontiers in Psychology* 3:224. doi: 10.3389/fpsyg.2012.00224.
- Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35, 606-621.
- Schwiedrzik, C., Singer, W., & Melloni, L. (2011). Subjective and objective learning effects dissociate in space and in time. *Proceedings of the National Academy of Sciences* 108, 4506–4511.
- Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition* 20(4), 1787–1792.
- Thiessen, E. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society of London: Biological Sciences* 372.
- Veenman, M. V., Elshout, J. J., & Meijer, J. (1997). The generality vs domainspecificity of metacognitive skills in novice learning across domains. *Learning and Instruction* 7(2), 187–209.
- Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia* 41, 108–114.
- Wang, T., & Saffran, J. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology* 5: 953
- Weiss, D. J., Gerfen, C., & Mitchel, A. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development* 5, 30–49.
- Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The neural basis of error-detection: conflict monitoring and the error-related negativity. *Psychological Review* 111, 931–959.
- Yim, D., & Rudoy, J. (2013). Implicit statistical learning and language skills in bilingual children. *Journal of Speech, Language and Hearing Research* 56, 310-322.
- Zarnoth, P. & Sniezek, J. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology* 33, 345–366.

FIGURE CAPTIONS

Figure 1. Descriptive data for experiment 1 (range of values, quartiles, median and individual data points marked by dots). **A.** Proportion of correct responses (hits and correct rejections) as a measure of accuracy. **B.** M-ratios as a measure of metacognitive sensitivity, which is $\text{meta-}d' / d'$, where $\text{meta-}d'$ measures sensitivity to the type-1 information available to make type-2 decisions. **C.** Mean confidence rating denoting metacognitive bias, or tendency of individuals to assign higher or lower confidence rating on each trial.

Figure 2. Descriptive data for experiment 2 (range of values, quartiles, median and individual data points marked by dots). **A.** Proportion of correct responses (hits and correct rejections) as a measure of accuracy. **B.** M-ratios as a measure of metacognitive sensitivity, which is $\text{meta-}d' / d'$, where $\text{meta-}d'$ measures sensitivity to the type-1 information available to make type-2 decisions. **C.** Mean confidence rating denoting metacognitive bias, or tendency of individuals to assign higher or lower confidence rating on each trial.

Figure 3. Random effects meta-analysis of effect of bilingualism on metacognitive efficiency. Upper line represents 95% confidence interval for the effect size in experiment 1, lower line represents 95% confidence interval for the effect size in experiment 2, and the squares show the estimated effect sizes. Diamond represents a pooled effect size, showing the 95% confidence interval and, vertically, the probability density of each value within this confidence interval.

Figure 4A (left). Prior and posterior probabilities for the difference in M-ratios between bilinguals and monolinguals in experiment 1.

Figure 4B (right). Prior and posterior probabilities for the difference in M-ratios between bilinguals and monolinguals in experiment 2.

Figure 5A (left). Confidence rating on correct responses and error trials. As neither the effect of the experiment, nor interaction between experiment and group were significant, we pooled the data on this figure across both experiments. Error bars indicate ± 2 standard errors around the mean.

Figure 5B (right). Prior and posterior probabilities for the difference in confidence on correct responses and confidence on error trials between monolinguals and bilinguals.