



Empirical investigation of the relationship between bilingualism and social flexibility

Marie-France Champoux-Larsson · Alexandra S. Dylman · Francisco Esteves

Received: 12 November 2020 / Revised: 4 March 2021 / Accepted: 5 March 2021
© The Author(s) 2021

Abstract Recently, a relationship between bilingualism and enhanced social flexibility has been suggested. However, research on the subject is scarce and what little exists is limited by several conceptual and methodological concerns. In the current study, we attempted to (a) replicate the findings from a study by Ikizer and Ramírez-Esparza (Bilingualism: Language and Cognition 21:957–969, 2018) by using the scales that the authors developed, and (b) test the concept of social flexibility experimentally with a switch-task using socially relevant stimuli. In the first part, participants ($n = 194$) filled out the scales developed by Ikizer and Ramírez-Esparza. We could not find that bilingualism leads to enhanced social flexibility. We did, however, find that higher level of education led to higher scores on the social flexibility scale. In the second part, a subsample ($n = 74$) from Part 1 completed a task where they were asked to identify the congruency between a face and a voice based on

either gender or emotion, and to switch between these two tasks. The experimental task did not show an advantage for the bilingual participants. On the contrary, higher proficiency in a second language led to lower accuracy in the congruent emotion condition, while level of education led to higher accuracy in that same condition. We suggest that factors other than bilingualism, such as level of education and biculturalism, most likely drove the effect found both in the current study and originally by Ikizer and Ramírez-Esparza.

Keywords Social flexibility · Bilingualism · Switch-task · Emotion

Empirical investigation of the relationship between bilingualism and social flexibility

The potential relationship between bilingualism and social interactions has not been extensively studied. Recently, however, results from a study by Ikizer and Ramírez-Esparza (2018) on bilingualism and social flexibility (where social flexibility was defined in terms of both acuity to relevant social cues and the ability to easily switch and adapt to different social environments) suggest that adult bilinguals are more socially flexible than adult monolinguals. In their study, Ikizer and Ramírez-Esparza developed and

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s41809-021-00076-7>.

M.-F. Champoux-Larsson (✉) · F. Esteves
Department of Psychology and Social Work, Mid Sweden University, 83125 Östersund, Sweden
e-mail: mfclarsson@gmail.com

A. S. Dylman
Department of Special Education, Stockholm University, 10691 Stockholm, Sweden

used a questionnaire that was based on the Trait Emotional Intelligence Questionnaire (TEIQue; Petrides, 2009) and found that bilinguals scored significantly higher on this social flexibility scale than monolinguals. Further, Ikizer and Ramírez-Esparza measured the frequency of social interactions and found that bilinguals had more social interactions than monolinguals, and that this effect was mediated by the social flexibility score. However, their investigation was correlational and contained a few conceptual and methodological concerns. Some of those concerns were pointed out by the authors themselves, while others were later raised in a commentary by Vives et al. (2018).

For instance, Ikizer and Ramírez-Esparza (2018) suggest that code-switching could be an important mechanism behind enhanced social flexibility. However, rather than measuring frequency of code-switching as such, the authors extrapolated the participants' code-switching frequency directly from their proficiency in a second language and from their frequency of use of a second language. Yet, code-switching does not necessarily overlap with proficiency and frequency of use since many bilinguals use their languages in different and separate environments (Grosjean, 2010). Additionally, the bilinguals in Ikizer and Ramírez-Esparza had a higher level of education than the monolingual participants, which was suggested by Vives et al. (2018) as being a probable confound influencing the results. Another issue was the exclusion of all monolingual participants who reported being bicultural. Although biculturalism and bilingualism are likely to strongly correlate, there are instances where one can be monolingual and still belong to two distinct cultures. This exclusion of bicultural monolinguals almost certainly led to a bilingual group that was more culturally diverse than the monolingual group. This makes it difficult to exclude biculturalism as yet another confound influencing the results.

Due in part to the abovementioned issues, Vives et al. (2018) argued that the effects reported by Ikizer and Ramírez-Esparza (2018) do not depend on bilingualism but rather are the result of other factors (i.e., cultural differences, educational level, biculturalism, and more). Although this is possible, Vives et al. presented no empirical data in their commentary to support their claim. We argue that such concerns, albeit justified, must be investigated more

systematically and that empirical evidence is necessary before completely rejecting the conclusions that were found in Ikizer and Ramírez-Esparza.

Indeed, there is evidence from other fields within psychology, such as emotion research, suggesting that language shapes social perception (e.g., Barrett et al., 2007; Nook et al., 2015). Whether having more than one language also shapes social perception differently is unknown, but the possibility remains. Furthermore, in a study by Marzecová et al. (2013), bilinguals outperformed monolinguals on a switch-task using social stimuli. There, participants performed a switch-task where they had to determine the gender or age of a face and to switch between the two. They found that bilinguals had a lower switch-cost in terms of reaction times for the gender trials, and greater accuracy in the switch and non-switch conditions. Although a bilingual advantage has been found earlier on switch-tasks (e.g., Costa et al., 2009; Jylkkä et al., 2017; Prior & MacWhinney, 2010; Stasenko et al., 2017, but see e.g., Hernández et al., 2013; Paap et al., 2017 for studies where no bilingual advantage was found), the particularity in Marzecová et al. was that they showed that the advantage encompasses social stimuli as well. Note however that the monolingual and bilingual participants in Marzecová et al. also most likely differed in terms of biculturalism. Namely, the bilingual group consisted uniquely of Hungary born Hungarian-Polish participants with a Polish mother and Hungarian father specifically, while the monolingual group consisted uniquely of Hungary born Hungarian monolinguals with no knowledge of another language. Given the stringent selection criterion for the bilingual group, it is highly likely that they not only were bilingual, but also bicultural. Therefore, given the scarcity of studies and incomplete evidence, there is not sufficient empirical support yet to confirm that a bilingual advantage in social flexibility exists. Additionally, given the proportions of the ongoing debate on the bilingual advantage in cognitive functions (see Lehtonen et al., 2018; Paap, 2019), it would be prudent and sound to first generate strong evidence for a potential social flexibility bilingual advantage before adding it to the discussion, particularly since social flexibility is not established in the literature yet.

Therefore, we designed a study where we explored the idea of a bilingual advantage in social flexibility in two steps. First, we attempted to replicate the findings from Ikizer and Ramírez-Esparza (2018) by using

their scales in a different bilingual population, and second, we tested the concept of social flexibility experimentally by using a switch-task similar to the one in Marzecová et al. (2013). Importantly, we addressed several of the issues raised above by including and controlling for other factors which may have affected the results in previous studies.

More specifically, we first asked a group of native Swedish speakers with varying levels of knowledge of a second language to fill out the social flexibility scale and the frequency of social interaction scale used in Ikizer and Ramírez-Esparza. The aim was to attempt to replicate the results where bilingualism led to higher scores on both the social flexibility scale and on the frequency of social interactions scale. Secondly, a subgroup of the participants from Part 1 performed a computerised task that was similar to the one used in Marzecová et al. (2013). With this task, we tested experimentally both acuity to relevant social cues (gender, emotion) and switching between tasks, which are the two facets of social flexibility as defined by Ikizer and Ramírez-Esparza (2018). Importantly, we wanted to increase the ecological validity of the task by combining stimuli that were both visual and auditory. Therefore, our adaptation of the social switch-task consisted of determining whether or not a voice and a face matched in terms of gender and of emotion, and to switch between those two factors (see “Methods” section for more information). Furthermore, while we used the same criterion as Ikizer and Ramírez-Esparza to define our bilinguals (i.e., frequency of use of a second language, and proficiency in the second language), we also collected information about code-switching. The aim was to test the suggestion by Ikizer and Ramírez-Esparza that frequent code-switching is a critical mechanism behind enhanced social flexibility. Finally, since level of education was pointed out as a possible confound by Vives et al., we also used this variable to test whether it would explain variation on the scales better than the bilingualism measurements could do.

If there is a relationship between bilingualism and social flexibility, we would expect to find, as in Ikizer and Ramírez-Esparza (2018), that higher bilingualism scores lead to higher scores on the scales. Furthermore, we would expect a higher frequency of code-switching to lead to higher scores on those scales if that is, as suggested, the mechanism behind the effect. On the other hand, if other factors such as level of

education drove the effect that was found earlier, we would expect the participants’ level of education to explain the variation on the scores. As for the experimental task, we predicted that if bilingualism does lead to enhanced social flexibility, the participant’s bilingualism in terms of second language proficiency, frequency of second language use, and/or frequency of code-switching would lead to better performance on the task. Namely, it would lead to higher accuracy, faster reaction times, and/or smaller switch-costs. However, if social flexibility is mostly driven by other factors, such as level of education, we expect level of education to be a better predictor of higher accuracy, shorter reaction times, and/or smaller switch-costs than the bilingualism measurements.

Methods

Part 1

Participants

Participants were recruited on campus and online via the University’s digital bulletin board and social media. A total of 194 participants partook in the first part of the study ($M_{age} = 37.5$, $SD_{age} = 13.2$, 74.7% females, 21.7% males, 3.6% other/unsure/preferred not to answer). Note that some of the participants ($n = 84$) participated on campus while the remaining of the sample participated online ($n = 110$). All participants reported having Swedish as a first language, and one or several additional languages. Observe that 91% of Swedes report knowing an additional language other than Swedish (European Commission, 2012), making it virtually impossible to find a group of Swedish monolinguals that would be large enough to be informative. The most frequent second language reported in our sample was English ($n = 119$), followed by German ($n = 19$), French ($n = 12$), regional Swedish dialects ($n = 10$), Norwegian ($n = 9$), Spanish ($n = 7$), Finnish ($n = 6$), Danish ($n = 2$), and finally, Bosnian, Greek, Italian, Mandarin, Polish, Romani, Swahili, Southern Sámi, Tamil, and Hungarian (all $n = 1$). Level of education (Mode = 4) was measured on a scale from 1 to 6 (1 = elementary school or lower: $n = 10$; 2 = high school: $n = 58$; 3 = professional education: $n = 27$; 4 = Bachelor’s degree: $n = 62$; 5 = Master’s degree:

$n = 30$; 6 = PhD: $n = 6$; one participant did not provide this information). Proficiency in Swedish was computed based on the mean of four scales from 1 to 10 asking participants about their skills compared to a native speaker for speaking, understanding speech, writing, and reading. Participants' average proficiency in Swedish was 9.9 ($SD = 0.5$). Participants also reported the frequency of use of Swedish for speaking, listening, writing and reading on a scale from 1 to 5 (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always). The mean of the four scales was computed to create a frequency of use score ($M = 4.6$, $SD = 0.7$). Proficiency in the most proficient second language and frequency of use of the most proficient second language were measured in the same way ($M_{proficiency} = 8.4$, $SD_{proficiency} = 1.5$; $M_{frequency} = 3.6$, $SD_{frequency} = 0.6$). One-tailed paired-samples t -tests revealed that participants were more proficient in Swedish than in their second language, $t(193) = 12.59$, $p < 0.001$, $d = 0.9$, and used Swedish more frequently than they used their second language, $t(193) = 18.11$, $p < 0.001$, $d = 1.3$. Finally, participants reported frequency of code-switching on scales from 1 to 5 (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always) by indicating how often they code-switch with parents, friends, and on social media respectively. The mean for the code-switching score, which consisted of the sum of all three scales, was 8.6 ($SD = 2.8$).

Materials

Authorised translators translated the Social Flexibility Scale, developed and validated by Ikizer and Ramírez-Esparza (2018), to Swedish and back-translated it to English. The scale consists of 11 items with different affirmations such as "I would describe myself as a flexible person" and "Generally, I'm able to adapt to new environments" where the participant indicates how much they agree with each statement on a scale from 1 to 7 (1 = completely disagree, 7 = completely agree). The items were presented in random order. The social flexibility score was computed by calculating the mean value of all items. Frequency of social interactions was measured with a modified version of the Frequency of Social Interactions scale (Ybarra et al., 2008) as was done in Ikizer and Ramírez-Esparza. Our adaptation of the scale, which included modern means of communication such as Skype and

chat services, consisted of questions asking about the frequency with which the participant (a) talked on the phone or via Skype, (b) texted or chatted, and (c) met, with family, other relatives, as well as friends and acquaintances respectively. The questions were answered on a scale from 1 to 6 (1 = never or rarely, 2 = at least once a year, 3 = at least once a month, 4 = at least once a week, 5 = daily, 6 = more than once daily). Here as well, the frequency of social interaction score consisted of the mean of all nine questions.

Analyses

Although Ikizer and Ramírez-Esparza (2018) conducted correlation analyses, they posit that it is bilingualism that leads to enhanced social flexibility and increased frequency of social interaction. In order to test this, we ran multiple linear regression analyses using various factors. Importantly however, based on several methodological reasons, we did not divide our participants into a group of monolinguals and a group of bilinguals as Ikizer and Ramírez-Esparza did. First, our sample did not contain participants reporting only speaking one language. Thus, all participants were at least somewhat bilingual, and creating a group of monolinguals and of bilinguals would not reflect our sample appropriately. Second, bilingualism is not a dichotomous variable (e.g., Bialystok, 2001; Kaushanskaya & Prior, 2015; Luk & Bialystok, 2013), and dividing a continuous variable into distinct groups increases the risk of Type I error (Cohen, 1983) and may thus inflate the rate of significant effects, leading to misleading results (MacCallum et al., 2002). Third, using bilingualism on a continuous scale has the advantage of considering the subtle variance between participants and how this correlates with various outcome variables rather than trying to see how groups differ on them, thus allowing for more fine-grained investigations of an effect (Luk & Bialystok, 2013). Nevertheless, in order to investigate the same factors used in Ikizer and Ramírez-Esparza, we used the same facets of bilingualism in our analyses as predictors. Namely, we used the participant's proficiency in their second language as well as the frequency of use of their second language as two separate continuous variables. Furthermore, in order to extend their results, we followed the premise in Ikizer and Ramírez-Esparza suggesting that code-switching could be the

mechanism behind the effects that they found. In order to do so, we used the frequency of code-switching as another continuous variable as predictor. Finally, to test the suggestion by Vives et al. (2018) that education most likely was behind the effect found in Ikizer and Ramírez-Esparza, we also used level of education as a predictor. We used those four predictors in two different analyses: one with the score of the social flexibility scale as the outcome variable, and one with frequency of social interactions as the outcome variable. Theoretically however, there was a high risk of multicollinearity between the predictors, particularly between the predictors that were a measurement of bilingualism. An examination of tolerance and the Variance Inflation Factors (VIF) revealed however that all values were satisfactory (all VIF < 1.59) and that multicollinearity was not a concern. All analyses were performed in Jasp (Jasp Team, 2019).

Part 2

Participants

Out of the participants from Part 1, a subset of 84 participants (i.e., those that participated in the study on campus) also partook in Part 2. However, three participants were excluded from the analyses for reporting not having normal or corrected sight, and two for reporting not having normal or corrected hearing. Furthermore, we controlled for outliers based on accuracy in the different conditions (congruent emotion trials, incongruent emotion trials, congruent gender trials, incongruent gender trials, see below for more information on the design). We excluded participants that had an accuracy below 3 standard deviations from the group mean in at least one of the conditions. This led to an exclusion of five more participants. Our sample for Part 2 thus consisted of 74 participants aged 18 to 68 years ($M = 36.5$ years, $SD = 12.7$; 32.4% males, 66.2% females, 1.4% other/unsure/preferred not to answer). The most frequent second language was English ($n = 61$), regional Swedish dialects ($n = 7$), German and Spanish (each $n = 2$), and finally Danish and Norwegian (each $n = 1$). Level of education (Mode = 2) was measured in Part 1 (1 = elementary school or lower: $n = 5$; 2 = high school: $n = 26$; 3 = professional education: $n = 10$; 4 = Bachelor's degree: $n = 21$; 5 = Master's degree: $n = 9$; 6 = PhD: $n = 3$). Proficiency in

Swedish ($M = 9.9$, $SD = 0.4$), frequency of use of Swedish ($M = 4.7$, $SD = 0.4$), proficiency in the second language ($M = 8$, $SD = 1.6$), frequency of use of the second language ($M = 3.5$, $SD = 0.7$), and frequency of code-switching ($M = 8.3$, $SD = 3.1$) were also measured in Part 1. One-tailed paired-samples t -tests revealed that participants in this subsample were more proficient in Swedish than in their second language, $t(73) = 9.84$, $p < 0.001$, $d = 1.14$, and used Swedish more frequently than they used their second language, $t(73) = 12.67$, $p < 0.001$, $d = 1.5$.

Stimuli

Visual stimuli The visual stimuli used in the experiment were selected from the Radboud Faces Database (RaFD: Langner et al., 2010). For this study, eight stimuli pictures (4 angry, 4 happy) portrayed by four different Caucasian actors (2 males, 2 females) were selected. The use of static pictures in the study of emotion presents several methodological issues (Russell, 1994). For instance, when depicting instances of emotions in stimuli, be it visually or orally, the represented emotion usually is a caricature of an instance of an emotion rather than a prototype (Goldstone et al., 2003), leading to lower ecological validity. We addressed this by using the genuineness score of the pictures (retrieved from Langner et al.'s validation data found in their supplementary materials) as the main selection criterion. Since the mean genuineness of anger pictures was significantly lower than the mean genuineness of happiness pictures ($t(76) = 9.84$, $p < 0.001$, $d = 2.23$), we selected the anger pictures with the highest genuineness scores and the corresponding happiness picture from the same actor in order to eliminate differences in genuineness between the two emotions. In our selected stimuli, there were no significant differences in genuineness based on the actor's gender (males: $M = 3.82$, $SD = 0.26$; females: $M = 3.79$, $SD = 0.19$; $F < 1$) or on the emotion (angry: $M = 3.71$, $SD = 0.25$; happy: $M = 3.9$, $SD = 0.12$; $F(1, 4) = 1.49$, $p = 0.289$, $\eta^2_p = 0.27$). The interaction between gender and emotion was not significant ($F(1, 4) = 1.19$, $p = 0.338$, $\eta^2_p = 0.23$).

Auditory stimuli The auditory stimuli were selected from the Juslin & Laukka, 2001 database (Juslin & Laukka, 2001; Laukka et al., 2005). Eight audio clips

where a semantically neutral utterance (“It is 11 o’clock”, in Swedish) expressed with an emotionally valenced prosody (4 angry, 4 happy) expressed by four different native Swedish speakers (2 males, 2 females) were selected. The recording of the utterances were available being expressed both with low and high intensity, but the utterances expressed at a lower intensity were scored as more natural (scores were obtained from a database provided by one of the authors: P. Laukka, personal communication, April 6, 2018). Following the same rationale as for the selection of the visual stimuli, the recordings with a lower intensity were selected in order to increase ecological validity. There were no significant differences in naturalness based on the actor’s gender (males: $M = 6.04$, $SD = 1.34$; females: $M = 6.04$, $SD = 0.37$) or the actor ($F < 1$) or on the emotion (angry: $M = 5.79$, $SD = 1.25$; happy: $M = 6.29$, $SD = 0.46$) that was uttered ($F < 1$). The interaction between gender and emotion was not significant ($F < 1$).

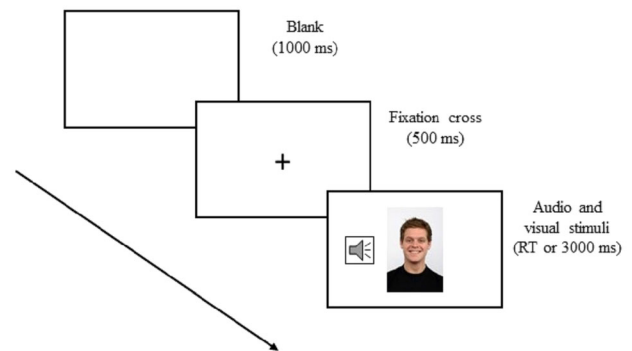
Design

Our task’s design was based on the study by Marzová et al. (2013) with a few modifications. As mentioned, in order to increase ecological validity, we chose to use both auditory and visual stimuli. In our task, the participants were instructed to determine whether the visual and auditory stimuli were congruent or incongruent based on either emotion or gender. These two tasks were presented in separate blocks in the first two blocks (i.e., one emotion block and one gender block) of the experiment (the order of presentation of the two blocks was counterbalanced across participants). These non-switch blocks (non-switch condition) allowed investigating social flexibility in terms of social cues acuity (which is one of the two facets in Ikizer and Ramírez-Esparza’s definition). For both non-switch blocks, each trial began with a blank background (1000 ms), followed by a fixation cross (500 ms). Afterwards, the face and voice were presented simultaneously, and the participant answered by pressing a key with the left finger on the keyboard if the visual and auditory stimuli were congruent (based on gender in the gender block, and based on emotion in the emotion block), and by pressing a key with the right finger on the keyboard if they were incongruent (the order of answer choice was

counterbalanced across participants). The next trial was presented after an answer was provided or after 3000 ms (see Fig. 1). Each non-switch block started with six practice trials (randomly selected from the eight possible face-voice combinations based on emotion and gender, see Table 1). Thereafter, 48 experimental trials were presented randomly. Half of the trials were congruent and half were incongruent.

After the first two blocks, a switch block was presented (switch condition) in order to measure participants’ social flexibility in terms of capacity to switch and adapt to different social tasks (which is the second of two facets in Ikizer and Ramírez-Esparza’s definition of social flexibility). In this block, participants had to respond based on emotion for half the trials and on gender for the other half. Each trial began with a blank background (500 ms), followed by a fixation cross (500 ms). Afterwards, a symbol indicating which criterion (i.e., emotion or gender) the evaluation should be based on appeared for 500 ms. Then, the face and voice were presented simultaneously, and the participant answered by pressing a key with the left finger on the keyboard if the visual and auditory stimuli were congruent, and by pressing a key with the right finger on the keyboard if they were incongruent (the order was also counterbalanced across participants, but was the same order that was used in the first two blocks). The next trial was presented after an answer was provided or after 3000 ms (see Fig. 2). The switch-block started with 12 practice trials (six emotion and six gender trials presented in a randomised order). Then, 96 experimental trials (48 emotion trials, 48 gender trials) were presented (again, in a randomised order). Half of the trials were congruent and half were incongruent.

In both the non-switch and switch conditions, the pairs of visual and auditory stimuli that were presented for a given trial were combined randomly. Furthermore, for the incongruent trials, only the target attribute was discrepant. For instance, for incongruent trials where congruency was to be judged based on emotion, the emotionality of the face and voice were different, but the gender of the face and voice were the same. Similarly, for incongruent trials where congruency was to be judged based on gender, the gender of the face and voice were different, but the emotion of the face and voice were the same (see Table 1). The same pictures and audio clips were used across all blocks. The pictures used in the practice trials were

Fig. 1 Non-switch block trial example*Non-Switch Block Trial Example***Table 1** Possible stimuli combinations for the different tasks

	Gender task		Emotion task	
	Visual stimuli	Auditory stimuli	Visual stimuli	Auditory stimuli
Congruent	Happy male	Happy male	Happy male	Happy male
	Angry male	Angry male	Angry male	Angry male
	Happy female	Happy female	Happy female	Happy female
	Angry female	Angry female	Angry female	Angry female
Discrepant	Happy male	Happy female	Happy male	Angry male
	Angry male	Angry female	Angry male	Happy male
	Happy female	Happy male	Happy female	Angry female
	Angry female	Angry male	Angry female	Happy female

portrayed by different actors and were not later used in the experimental trials. As for the sound clips used in practice trials, a question version of the utterance (“Is it 11 o’clock?”, in Swedish) was used for the auditory stimuli, and was not later used in the experimental trials.

Procedure

After filling out the questionnaire used in Part 1, the experiment (programmed and presented in E-Prime Version 2.0: Psychology Software Tools, 2012) began. The experimenter explained the procedure and specific instructions for each block were presented in writing during the experiment. Participants were compensated for their participation in the experiment with a gift certificate for a movie ticket (i.e., only participants in Part 2 received compensation). The study followed all relevant ethical regulations and Swedish laws concerning research with human participants (for both Part 1 and Part 2).

Analyses

The same rationale as in Part 1 was followed to choose the predictors in Part 2. Namely, proficiency in the second language, frequency of use of the second language, frequency of code-switching, and level of education were used as continuous independent variables. Furthermore, since age is related to a decrease in cognitive functions (e.g., Kray & Lindenberger, 2000; Tun & Lachman, 2008) and ability to identify emotions (Ruffman et al., 2008), we added the participants’ age as a predictor. This was particularly important given the wide age range of our sample.

For the experimental task, only correct answers were included in the analysis of reaction times. Furthermore, correct answers with a reaction time faster than 700 ms were treated as errors given that participants cannot have had time to process the auditory stimuli and make a decision that were faster than that (Pell & Kotz, 2011; Rigoulot et al., 2013). To investigate social cues acuity, accuracy and reaction times for the congruent and incongruent trials in the

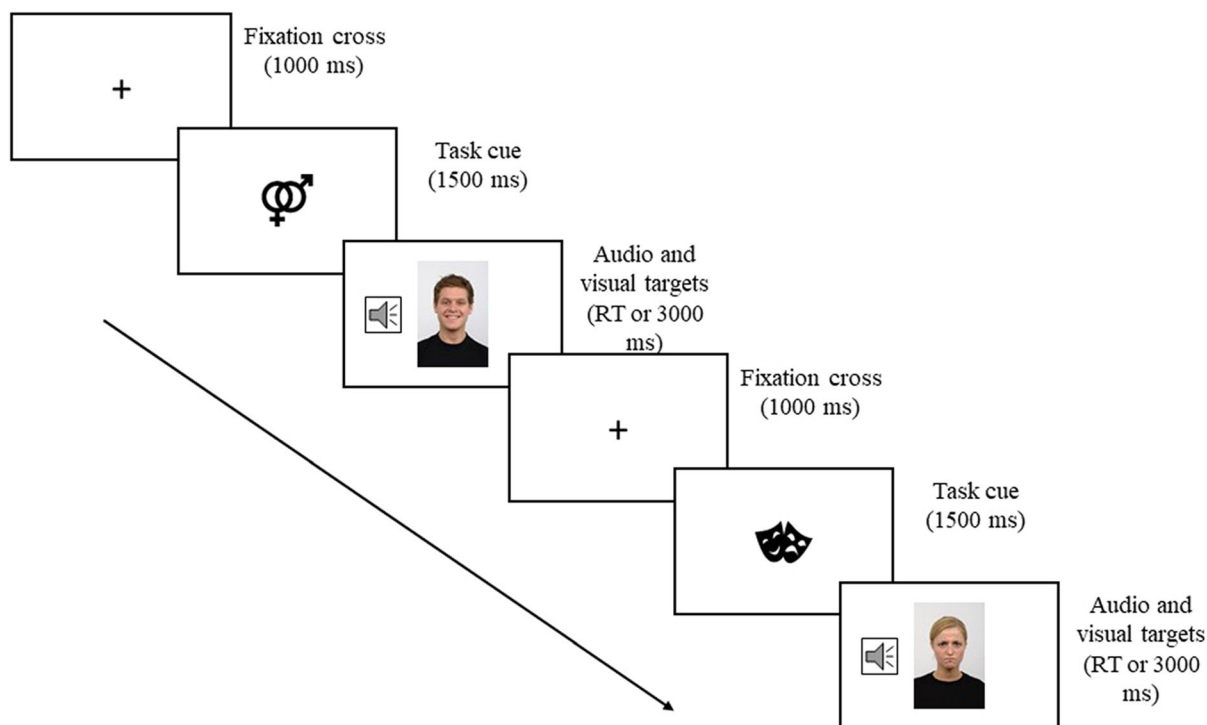


Fig. 2 Switch block trials example

emotion and gender non-switch blocks were analysed individually. Furthermore, switch-cost was investigated as well. A switch-cost is represented by slower reaction times when responding to a trial based on criterion A, when the preceding trial had to be responded to based on criterion B (and vice versa). Switch-costs were thus calculated by computing the difference in mean reaction times in switch trials compared to non-switch trials (i.e. gender trials preceded by emotion trials compared to gender trials preceded by gender trials, and emotion trials preceded by gender trials compared to emotion trials preceded by emotion trials).

Multiple regression analyses were conducted with the variables presented above as predictors. Theoretically however, there was a high risk of multicollinearity between the predictors here as well. Collinearity was thus controlled for by examining tolerance and the Variance Inflation Factors (VIF), which were satisfactory (all VIF < 1.7). All analyses were conducted in Jasp (Jasp Team, 2019).

Results

Part 1

The multiple linear regression with second language proficiency, second language frequency of use, frequency of code-switching, and education as predictors and the social flexibility score ($M = 5.03$, $SD = 0.89$) as outcome variable was not significant but approached significance, $F(4, 188) = 2.24$, $p = 0.066$, $R^2_{adj} = 0.025$. Because the model approached significance, we took a closer look at the predictors. Second language proficiency, second language frequency of use, and frequency of code-switching were not significant, but education was ($\beta = 0.12$, $p = 0.025$). For exploratory purposes only, we ran a simple linear regression analysis with education as a predictor and the social flexibility score as the outcome variable. This model was significant, $F(1, 191) = 5.64$, $p = 0.019$, $\beta = 0.12$, $R^2 = 0.029$, suggesting that a model without the bilingualism measurements better predicts social flexibility. Note however that a post hoc power analysis showed that the achieved power was

Table 2 Descriptive data for predictors and outcome variables in Part 1

	<i>M</i> or mode	<i>SD</i>	Min	Max	Possible range
Predictors					
Second language proficiency	8.4	1.5	1.7	10	0–10
Second language use	3.6	0.6	2	5	1–5
Code-switching	8.6	2.8	3	15	3–20
Level of education	4	–	1	6	1–6
Outcome					
Social flexibility	5.03	0.89	2.45	6.73	1–7
Frequency of social interactions	3.28	0.62	1.67	4.92	1–6

relatively low (0.67), suggesting that this result should be interpreted carefully. As for the multiple linear regression with second language proficiency, second language frequency of use, frequency of code-switching, and education as predictors and the frequency of social interactions score as outcome variable ($M = 3.28$, $SD = 0.62$), it was not significant, $F(4, 188) = 0.75$, $p = 0.56$, $R^2_{adj} = -0.005$. Please see Table 2 for a summary of the predictors and outcome variables, and the *Supplementary Information* online for histograms and scatter plots over the predictors and outcome variables.

Part 2

The regression model for the accuracy on the congruent emotion trials ($M = 19.5$ correct answers, $SD = 3.5$) was significant, $F(5, 68) = 3.29$, $p = 0.01$, $R^2_{adj} = 0.14$, where higher proficiency led to lower accuracy ($\beta = -0.43$, $p = 0.004$), and higher education predicted higher accuracy ($\beta = 0.39$, $p = 0.003$). A post hoc power analysis indicated that the achieved power of this analysis was 0.76, which is near the minimum threshold of 0.8. However, the regression models for accuracy on incongruent emotion trials ($M = 16.9$ correct answers, $SD = 3.2$), $F(5, 68) = 1.34$, $p = 0.26$, congruent gender trials ($M = 22.9$, $SD = 1.3$), $F(5, 68) = 1.44$, $p = 0.22$, and incongruent gender trials ($M = 23$ correct answers, $SD = 1.2$), $F(5, 68) = 1.44$, $p = 0.22$, were all non-significant.

To continue, the regression analyses using the same predictors were performed but with reaction times as the outcome variable. The model for congruent emotion trials ($M = 1970$ ms, $SD = 279$ ms) was not significant, $F(5, 68) = 1.43$, $p = 0.23$, nor was the model for incongruent emotion trials ($M = 2026$ ms,

$SD = 238$ ms), $F(5, 68) = 1.67$, $p = 0.153$. As for gender, the models for the congruent trials ($M = 1444$ ms, $SD = 302$ ms), $F(5, 68) = 1.57$, $p = 0.18$, and for incongruent trials ($M = 1421$ ms, $SD = 281$ ms), $F(5, 65) = 2.72$, $p = 0.057$ were not significant. Since the model for incongruent gender trials approached significance however, we took a closer look at the factors. None were significant except age ($p = 0.016$). For exploratory purposes only, we ran a simple linear regression analysis with age as predictor and reaction times on incongruent gender trials as outcome variable. The model was significant, $F(1, 72) = 8.84$, $p = 0.004$, $R^2 = 0.097$, with a post hoc power analysis showing that the achieved power was 0.84, suggesting that age ($\beta = 0.33$) was a better predictor of performance on incongruent gender trials than other factors, with participants being slower as they were older.

In order to investigate switching, the three bilingualism scores, level of education, and age were used as predictors in multivariate linear regressions analysis with switch-cost for emotion to gender trials ($M = -103.2$, $SD = 131$) and switch-cost for the gender to emotion trials ($M = -25.6$, $SD = 140$) as the outcome variables. The models were non-significant both for emotion to gender, $F(5, 68) = 0.87$, $p = 0.504$ and for gender to emotion, $F(5, 68) = 0.96$, $p = 0.45$ (see Table 3 for raw reaction times, Table 4 for a summary of the predictors and outcome variables, and the *Supplementary Information* online for a visualisation of the distribution of the predictors and outcome variables and significant results).

We also investigated whether participants were learning the task in terms of faster reaction times in the different blocks. To do so, we divided the emotion, gender, and switch blocks into four time periods and calculated the mean reaction time for each participant

Table 3 Raw reaction times for the switch block

	<i>M</i>	<i>SD</i>	Min	Max
Non-switch trials	1584	284	1039	2331
Non-switch emotion trials	1801	321	1193	2399
Non-switch gender trials	1404	281	850	2286
Switch trials	1651	289	1085	2332
Switch trials from emotion to gender	1508	306	988	2315
Switch trials from gender to emotion	1827	299	1127	2431

Table 4 Descriptive data for predictors and outcome variables in Part 2

	<i>M</i> or mode	<i>SD</i>	Min	Max	Possible range
Second language proficiency	8	1.6	3.7	10	1–10
Second language frequency	3.5	0.7	2	5	1–5
Code-switching	8.3	3.1	3	15	3–20
Level of education	2	–	1	6	1–6
Age	36.5	12.7	18	68	18+
Accuracy congruent emotion	19.5	3.5	6	24	0–24
Accuracy incongruent emotion	16.9	3.2	9	23	0–24
Accuracy congruent gender	22.9	1.3	18	24	0–24
Accuracy incongruent gender	23	1.2	19	24	0–24
RT (ms) congruent emotion	1970	279	1330	2502	700–3000
RT (ms) incongruent emotion	2026	238	1410	2586	700–3000
RT (ms) congruent gender	1444	302	891	2074	700–3000
RT (ms) incongruent gender	1421	281	877	2198	700–3000
Switch-cost to gender (ms)	– 103.2	131	– 428	329	n/a
Switch-cost to emotion (ms)	– 25.6	140	– 342	250	n/a

in each time period. We then conducted a repeated measures ANOVA for each block. For the emotion block, there was a significant effect of the time period, $F(3, 219) = 18.76$, $p < 0.001$, $\eta^2 = 0.2$. Post-hoc comparisons with Bonferroni correction showed that reaction times in the first time period ($M = 2080$, $SD = 268$) were equal to the reaction times in the second time period ($M = 2042$, $SD = 280$), but longer than reaction times in the third ($M = 1952$, $SD = 274$) and fourth periods ($M = 1924$, $SD = 302$, both $ps < 0.001$). Furthermore, the reaction times in the second time period were significantly longer than those of the third and fourth time periods (both $ps < 0.001$). There was no difference between the third and fourth time periods. As for the gender block, the Mauchly's test of sphericity was significant and a Greenhouse–Geisser correction was used. However, non-adjusted degrees of freedom are reported for increased readability. There was a significant effect of time period, $F(3, 219) = 22.15$, $p < 0.001$, $\eta^2 = 0.23$. Post-hoc

comparisons with Bonferroni correction showed that reaction times in the first time period ($M = 1531$, $SD = 339$) were longer than reaction times in the second time period ($M = 1460$, $SD = 315$, $p = 0.02$), and longer than reaction times in the third ($M = 1396$, $SD = 307$) and fourth time periods ($M = 1346$, $SD = 289$, both $ps < 0.001$). The reaction times in the second time period were longer than reaction times in the third ($p = 0.05$) and fourth ($p < 0.001$) time periods. There was no difference between the third and fourth time periods. Finally, for the switch block, the Mauchly's test of sphericity was significant as well and a Greenhouse–Geisser correction was used (non-adjusted degrees of freedom are reported). There was a significant effect of time period, $F(3, 219) = 6.36$, $p < 0.001$, $\eta^2 = 0.08$. Post-hoc comparisons with Bonferroni correction showed that the reaction times during the first time period ($M = 1698$, $SD = 318$) were longer than reaction times in the second time period ($M = 1635$, $SD = 314$, $p = 0.04$), and longer

than reaction times in the third ($M = 1610$, $SD = 307$) and fourth time periods ($M = 1611$, $SD = 287$, both $ps = 0.001$). There were no significant differences between the other time periods. These analyses show that, as would be expected, some degree of learning occurred in all blocks, but that it plateaued in all blocks.

Discussion

In this study, we investigated social flexibility as a function of bilingualism in order to further examine the findings of Ikizer and Ramírez-Esparza (2018), who found bilingualism to correlate with larger social flexibility and frequency of social interactions. We first attempted to replicate their findings by using their scales in a different population of bilinguals. Our results suggest that level of education is a better predictor of social flexibility than any of the bilingualism measurements that we used. We also attempted to test the concept of social flexibility behaviourally by asking a sub-sample to complete a switch-task using socially relevant stimuli. Here as well, our results suggest that level of education is a better predictor of social flexibility (assuming that this is what the switch-task was measuring), and that higher proficiency actually leads to lower accuracy.

Thus, based on our results, we cannot support the hypothesis that bilinguals are more socially flexible than monolinguals, at least not in our sample. Regardless of how bilingualism was defined (based on proficiency in the second language, on frequency of use of the second language, or on the frequency of language switching), bilingualism could not predict higher scores on the social flexibility scale nor on the frequency of social interaction scales. Neither could bilingualism predict a better performance on any of the aspects of the task in our experiment (accuracy, reaction times, switch-cost). Furthermore, the hypothesis that code-switching is an underlying mechanism that could lead to enhanced social flexibility was not confirmed. This is in line with new research by Jylkkä et al. (2020) showing that frequent code-switching in daily life is associated with lower monitoring skills. Based on this, a positive effect of frequency of code-switching on social flexibility would actually have been surprising.

In fact, when investigating task accuracy in the non-switch condition, level of proficiency in a second language significantly predicted a *lower* accuracy in the emotion block for the congruent trials. Although we cannot explain why higher proficiency in a second language led to lower accuracy in emotional cue acuity, constructionist theories of emotion, which posit that language shapes the perception of emotions (e.g., Barrett, 2017; Barrett et al., 2007; Lindquist, 2017; Lindquist et al., 2015) offer a theoretical framework with which we could attempt an explanation. Based on this theoretical framework, our results suggest that having more than one language impedes the interpretation of emotional cues, at least when they are presented out of context as in the current study. Indeed, it has been shown that having strict and categorical, “black and white” concepts of emotions reduces the threshold at which facial movements are identified as a specific facial expression representing an emotion (Satpute et al., 2016). On the other hand, having less rigid and more fluid concepts of emotion requires more contextual information before an emotion can be inferred from facial movements (Satpute et al., 2016). Furthermore, since the essence of emotion words is internally represented differently across languages (e.g., Altarriba, 2003), it is possible that the proficient bilingual has more nuanced or ambiguous concepts of emotions due to those conceptual differences across languages. If so, it is likely that they would have less categorical concepts of emotions, thus requiring more contextual information before they can interpret a facial movement or voice modulation as a specific emotion. Since our task did not provide any contextual information, we hypothesise that it was more difficult for participants with more fluid concepts of emotion, as perhaps the most fluent of our bilinguals were, to be accurate. The lack of effect of proficiency in a second language on the incongruent trials could be explained by the fact that incongruent trials are more difficult for all participants, even those with more categorical concepts of emotions. Of course, this is only a tentative explanation and will need to be investigated more carefully.

Additionally, level of education did predict higher scores on the social flexibility scale and better accuracy on congruent emotion trials in the current study (but did not predict any other variable), suggesting that level of education may have contributed to the effect that Ikizer and Ramírez-Esparza

found. Indeed, in their study, the bilingual group was more educated than the monolingual group. Note however that the better performance of those who were more educated can be related to better performance due to higher executive functions in our task. While our task aimed to operationalise social flexibility experimentally by using socially relevant stimuli, it could be that what we observe is merely better executive functions regardless of social flexibility. Related to executive functions, we found that the older the participants were, the slower they were on incongruent gender trials. Why age had an effect on incongruent gender trials only is not clear, but the effect that age had on reaction times is consistent with previous findings on an age-related decline on speed for cognitive tasks (e.g., Deary & Der, 2005; Thompson et al., 2014).

Another aspect which might have affected the results relates to the issue of biculturalism. The bilingual group in the original study (and interestingly, in Marzecová et al.'s, 2013, as well) was highly likely to be bicultural while the monolingual group was, according to the authors, monocultural (Ikizer & Ramírez-Esparza, 2018). Meanwhile, the sample in the current study was likely to be more culturally homogenous than the sample that Ikizer and Ramírez-Esparza (2018) tested. Although we did not measure biculturalism per se, our participants arguably all came from a more homogenous cultural background, with all of them having Swedish as a first language and most of them having English as a second language. This might be a notable difference between the populations which may have led to different findings, even when using the same outcome variables. Indeed, Ikizer and Ramírez-Esparza suggested that alternating between two *cultural* worlds may lead to more social flexibility. However, one can be bicultural without being bilingual, and vice versa (Grosjean, 2015). We suggest that, on a conceptual level, the resemblance between cultural switching and social environment switching is larger than the relationship between language switching and social switching. For instance, biculturals are characterised as being active in two different cultures, adapting various aspects of their lives (such as beliefs, norms and values), and combining different aspects of their two cultures (Grosjean, 2010; Nguyen & Benet-Martínez, 2007). The similarities between the two concepts makes it plausible that biculturalism can contribute to a higher

degree of social flexibility. However, this should be studied more closely before such speculative interpretations can be established.

Also, it is worth noting that our population of monolinguals and bilinguals differed from the population that was studied in Ikizer and Ramírez-Esparza (2018) on another aspect, in that there were no monolinguals in the current study. If our population was indeed more bilingual and there was not enough variance when it comes to their language profile, this could explain why we did not find an effect of bilingualism on social flexibility. However, second language proficiency did have an effect when it came to accuracy on congruent emotion trials, suggesting that the language profile of our sample was varied enough to detect at least some potential effects of bilingualism. Furthermore, we used bilingualism as a continuous variable specifically in order to be able to detect more fine-grained differences. However, bilingualism is a complex concept consisting of several elements, which is partially illustrated by the different facets of bilingualism that were measured in this study. Future research should thus address how these various facets of bilingualism can modulate the effects that we found when it comes to proficiency. Indeed, a recent study by Champoux-Larsson and Dylman (2021) shows that the operationalisation of bilingualism may affect the significance of the results on cognitive tasks. Ideally, future research should be conducted in populations where monolinguals who have no or very little knowledge of a second language can be recruited. By including a larger number of monolinguals, a broader understanding could be gained. Furthermore, a larger number of monolinguals would allow treating bilingualism as a dichotomous variable if one would prefer to adopt this more traditional approach (but see Champoux-Larsson & Dylman, 2021, for the consequences of dichotomising a variable such as bilingualism on the statistical significance of results).

It is also important to point out that both Ikizer and Ramírez-Esparza (2018) and the current study only used self-reported measurements of second language proficiency. Self-reported measurements do not necessarily correlate strongly with objective measurements, at least for dominance (e.g., Gollan et al., 2011; Sheng et al., 2014). On the other hand, since the second languages of our participants varied greatly, objectively measuring their proficiency in their second

language would have been methodologically impractical for languages where no objective tests exist. Furthermore, the heterogeneous profiles of our participants increase the generalisability of our results to a larger population, and consequently, our conclusions do not need to be limited to a particular type of bilinguals with specific first and second languages. Nonetheless, future research should address the matter of proficiency measurements by objectively assessing this variable instead of using self-reported measures only, or in combination with them.

Moreover, using emotion and gender as social cues by mixing visual and auditory mediums is not without its challenges. As illustrated by the low switch-cost for the emotion trials, the design of this study using vocally expressed emotions may have led to a floor effect, thus minimizing the likelihood of observing a switch-cost for the emotion trials. Emotions expressed vocally require a relatively long time to be recognized. For instance Goerlich et al. (2012) found that, although we are quite accurate when identifying emotions in speech, emotional prosody may still take 600 to 700 ms to process. Note however that Goerlich et al. used auditory stimuli that was rated as highly positive or highly negative. When we chose our auditory stimuli, we prioritised stimuli with lower intensity since they were perceived as more natural. The lower intensity of our stimuli may have made them harder to identify, and/or may have required even more time. To add to this, different emotions are processed with different speeds, and certain emotions are recognized faster than others (Pell & Kotz, 2011; Rigoulot et al., 2013). For instance, vocally expressed happiness takes significantly longer to be recognised than vocally expressed anger, and even anger, which is more hastily recognized than happiness, can still require around 700 ms to be identified in utterances (Pell & Kotz, 2011; Rigoulot et al., 2013). Due to the relatively slow temporal course of emotion recognition in auditory stimuli, potential switch-costs might have been too small to have had an observable effect on the reaction times in the current study. Future studies may want to modify the recognition task in order to overcome this issue, for example by presenting the stimuli within the same modality.

Although the current paper did not originally aim to look specifically into age as a factor, given that age affects both cognitive processing (e.g., Kray & Lindenberger, 2000; Tun & Lachman, 2008) and

emotion perception (Ruffman et al., 2008), this would be interesting to investigate in the future. However, as the aim of this paper was to investigate social flexibility and bilingualism, not cognitive processing and emotion perception as a function of age, there is not enough variability in our sample as far as age is concerned to allow comparisons between, say younger and older adults. However, future studies may want to investigate this more closely.

Additionally, there is a large body of research showing that emotions are perceived less intensely in a second language (e.g., Caldwell-Harris, 2014, 2015; Dylman & Bjärtå, 2018; Harris et al., 2003; Pavlenko, 2005; Puntoni et al., 2009) and it has been suggested that lower emotionality in a second language may explain why we tend to make different decisions in a second language compared to a first language (a phenomenon tokened as the Foreign Language effect, e.g., Cipolletti et al., 2016; Corey et al., 2017; Costa et al., 2014; Dylman & Champoux-Larsson, 2020; Geipel et al., 2015; Hayakawa et al., 2016; Keysar et al., 2012). Here, participants were only tested in their first language. However, further research should address the perception of emotions in a second language context.

Furthermore, the relationship between emotion and gender in relation to social flexibility is a relevant topic that should be investigated in more detail. Indeed, there are mixed results showing that gender differences in emotion perception and memory may vary as a function of several factors such as the gender of participant, of the person expressing the emotion and/or of the emotion being expressed (e.g., Cortes et al., 2017; Franklin & Adams, 2010; Gupta & Srinivasan, 2009; Krumhuber & Manstead, 2011). However, since this study was not designed to investigate the intricate relationship between these factors and social flexibility, but instead aimed to specifically investigate the relationship between bilingualism and social flexibility, we suggest that further research should address the former.

Finally, on a cautious note, we would like to point out again that social flexibility is not an established concept in the literature. In this study, we chose to use the definition proposed by Ikizer and Ramírez-Esparza (2018) as a starting point. However, particularly in the light of our inconsistent results, before testing whether or not bilinguals are more socially flexible, it would be methodologically sound to first establish what social

flexibility is, what it consists of, and how it can be operationalised and measured.

Nonetheless, our study contributes to the body of research on the various effects that bilingualism can have on other non-linguistic processes by providing more knowledge on the social life of bilinguals, an area that has received little attention to date. Our results concurrently provide empirical evidence to justify some of the concerns raised by Vives et al. (2018) and raises questions on the effects that were originally found in Ikizer and Ramírez-Esparza (2018). Our study also presents new questions when it comes to the effect of bilingualism on the inference of emotion based on visual and auditory cues presented in a context-less paradigm. Namely, we found that a higher level of proficiency in a second language led to lower accuracy in emotion perception. This result, which was somewhat surprising to us, deserves to be investigated further.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Marie-France Champoux-Larsson. The first draft of the manuscript was written by Marie-France Champoux-Larsson and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Mid Sweden University.

Declarations

Conflicts of interest The authors have no conflicts of interest to report.

Consent to participate Written informed consent was obtained from participants.

Ethics approval All national and international laws and regulation were followed.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly

from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altarriba, J. (2003). Does *cariño* equal 'liking'? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism*, 7, 305–322. <https://doi.org/10.1177/13670069030070030501>
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23. <https://doi.org/10.1093/scan/nsw154>
- Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *TRENDS in Cognitive Science*, 11, 327–332. <https://doi.org/10.1016/j.tics.2007.06.003>
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, & cognition*. Cambridge University Press.
- Caldwell-Harris, C. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.01055>
- Caldwell-Harris, C. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science*, 24(3), 214–219. <https://doi.org/10.1177/0963721414566268>
- Champoux-Larsson, M.-F., & Dylman, A. S. (2021). Different measurements of bilingualism and their effect on performance on a Simon task. *Applied Psycholinguistics*, 42(2), 505–526. <https://doi.org/10.1017/S0142716420000661>
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral Foreign-language effect. *Philosophical Psychology*, 29(1), 23–40. <https://doi.org/10.1080/09515089.2014.993063>
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–253. <https://doi.org/10.1177/014662168300700301>
- Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 1109–1128. <https://doi.org/10.1037/xlm0000356>
- Cortes, D. S., Laukka, P., Lindahl, C., & Fischer, H. (2017). Memory for faces and voices varies as a function of sex and expressed emotion. *PLoS ONE*, 12, 1–19. <https://doi.org/10.1371/journal.pone.0178423>
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., Keysar, B., & Sigman, M. (2014). Your morals depend on language. *PLoS ONE*, 9, e94842. <https://doi.org/10.1371/journal.pone.0094842>
- Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, 113(2), 135–149. <https://doi.org/10.1016/j.cognition.2009.08.001>
- Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Ageing*,

- Neuropsychology, and Cognition*, 12, 187–215. <https://doi.org/10.1080/13825580590969235>
- Dylman, A. S., & Bjärtå, A. (2018). When your heart is in your mouth: The effect of second language use on negative emotions. *Cognition and Emotion*, 33(6), 1284–1290. <https://doi.org/10.1080/02699931.2018.1540403>
- Dylman, A. S., & Champoux-Larsson, M.-F. (2020). It's (not) all Greek to me: Boundaries of the foreign language effect. *Cognition*, 196, 104148. <https://doi.org/10.1016/j.cognition.2019.104148>
- European Commission. (2012). *Europeans and their languages* (Special Eurobarometer 386). https://ec.europa.eu/comfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf
- Franklin, R. G., Jr., & Adams, R. B., Jr. (2010). What makes a face memorable? The relationship between face memory and emotional state reasoning. *Personality and Individual Differences*, 49, 8–12. <https://doi.org/10.1016/j.paid.2010.02.031>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, 59, 8–17. <https://doi.org/10.1016/j.jesp.2015.02.001>
- Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience*, 24(8), 1725–1741. https://doi.org/10.1162/jocn_a_00213
- Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, 31, 169–180. <https://doi.org/10.3758/bf03194377>
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R., & Cera, C. M. (2011). Self-ratings of spoken language dominance: A multilingual naming test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 15, 594–615. <https://doi.org/10.1037/0893-3200.10.3.243>
- Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.
- Grosjean, F. (2015). Bicultural bilinguals. *International Journal of Bilingualism*, 19, 572–586. <https://doi.org/10.1177/1367006914526297>
- Gupta, R., & Srinivasan, N. (2009). Emotions help memory for faces: Role of whole and parts. *Cognition and Emotion*, 23, 807–816. <https://doi.org/10.1080/02699930802193425>
- Harris, C., Ayçiçeği, A., & Gleason, J. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24, 561–579. <https://doi.org/10.1017/S0142716403000286>
- Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*, 20(11), 791–793. <https://doi.org/10.1016/j.tics.2016.08.004>
- Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, 69(3), 257–276. <https://doi.org/10.1016/j.jml.2013.06.004>
- Ikizer, E., & Ramírez-Esparza, N. (2018). Bilinguals' social flexibility. *Bilingualism: Language and Cognition*, 21, 957–969. <https://doi.org/10.1017/S1366728917000414>
- JASP Team. (2019). JASP (version 0.14) [Computer software]. <https://jasp-stats.org/>
- Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, 1, 381–412. <https://doi.org/10.1037/1528-3542.1.4.381>
- Jylkkä, J., Laine, M., & Lehtonen, M. (2020). Does language switching behavior rely on general executive functions? *Bilingualism: Language and Cognition*. <https://doi.org/10.1017/S1366728920000619>
- Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals: An Internet-based study. *Journal of Cognitive Psychology*, 29, 404–419. <https://doi.org/10.1080/20445911.2017.1282489>
- Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition*, 18, 27–28. <https://doi.org/10.1017/S1366728914000510>
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23, 661–668. <https://doi.org/10.1177/0956797611432178>
- Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, 15(1), 126–147. <https://doi.org/10.1037/0882-7974.15.1.126>
- Krumhuber, E. G., & Manstead, A. S. R. (2011). When memory is better for out-group faces: On negative emotions and gender roles. *Journal of Nonverbal Behavior*, 35, 51–61. <https://doi.org/10.1007/s10919-010-0096-8>
- Langner, O., Dotsch, R., Bijlstra, G., Widboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24, 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, 19, 633–653. <https://doi.org/10.1080/02699930441000445>
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, 144, 394–425. <https://doi.org/10.1037/bul0000142>
- Lindquist, K. A. (2017). The role of language in emotion: Existing evidence and future directions. *Current Opinion in Psychology*, 17, 135–139. <https://doi.org/10.1016/j.copsyc.2017.07.006>
- Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, 24(2), 99–108. <https://doi.org/10.1177/0963721414553440>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25, 605–621. <https://doi.org/10.1080/20445911.2013.795574>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative

- variables. *Psychological Methods*, 7, 19–40. <https://doi.org/10.1037/1082-989X.7.1.19>
- Marzecová, A., Bukowski, M., Correa, Á., Boros, M., Lupiáñez, J., & Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: The role of flexibility in temporal preparation and category switching. *Journal of Cognitive Psychology*, 25, 586–604. <https://doi.org/10.1080/20445911.2013.809348>
- Nguyen, A.-M.D., & Benet-Martínez, V. (2007). Biculturalism unpacked: Components, measurement, individual differences, and outcomes. *Social and Personality Psychology Compass*, 1, 101–114. <https://doi.org/10.1111/j.1751-9004.2007.00029.x>
- Nook, E. C., Linquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speeds and shapes facial emotion recognition. *Emotion*, 15(5), 569–578. <https://doi.org/10.1037/a0039166>
- Paap, K. R. (2019). The bilingual advantage debate: Quantity and quality of the evidence. In J. W. Schwieter (Ed.), *The handbook of neuroscience of multilingualism* (pp. 701–735). Wiley-Blackwell.
- Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinskay, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, 29(2), 89–112. <https://doi.org/10.1080/20445911.2016.12488436>
- Pavlenko, A. (2005). *Emotions and multilingualism*. Cambridge University Press.
- Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, 6, 1–16. <https://doi.org/10.1371/journal.pone.0027256>
- Petrides, K. V. (2009). Psychometric properties of the Trait Emotional Intelligence Questionnaire (TEIQue). In J. D. A. Parker, D. H. Saklofske, & C. Stough (Eds.), *Assessing emotional intelligence* (pp. 85–101). Springer.
- Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253–262. <https://doi.org/10.1017/S1366728909990526>
- Psychology Software Tools. (2012). E-Prime [Version 2.0]. <https://www.pstnet.com>
- Puntoni, S., De Langhe, B., & Van Osselaer, S. M. J. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of Consumer Research*, 35, 1012–1025. <https://doi.org/10.1086/595022>
- Rigoulot, S., Wassiliwizky, E., & Pell, M. D. (2013). Feeling backwards? How temporal order in speech affects the time course of vocal emotion recognition. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2013.00367>
- Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, 32, 863–881. <https://doi.org/10.1016/j.neubiorev.2008.01.001>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141. <https://doi.org/10.1037/0033-2909.115.1.102>
- Satpute, A. B., Nook, E., Narayanan, S., Shu, J., Weber, J., & Ochsner, K. N. (2016). Emotions in “black and white” or shades of gray? How we think about emotion shapes our perception and neural representation of emotion. *Psychological Science*, 27, 1428–1442. <https://doi.org/10.1177/0956797616661555>
- Sheng, L., Lu, Y., & Gollan, T. H. (2014). Assessing language dominance in Mandarin–English bilinguals: Convergence and divergence between subjective and objective measures. *Bilingualism: Language and Cognition*, 17, 364–383. <https://doi.org/10.1017/S1366728913000424>
- Stasenko, A., Matt, G. E., & Gollan, T. H. (2017). A relative bilingual advantage in switching with preparation: Nuanced explorations of the proposed association between bilingualism and task switching. *Journal of Experimental Psychology: General*, 146(11), 1527–1550. <https://doi.org/10.1037/xge0000340>
- Thompson, J. J., Blair, M. R., & Henrey, A. J. (2014). Over the hill at 24: Persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. *PLoS ONE*, 9(4), e94215. <https://doi.org/10.1371/journal.pone.0094215>
- Tun, P. A., & Lachman, M. E. (2008). Age differences in reaction time and attention in a national telephone sample of adults: Education, sex, and task complexity matter. *Developmental Psychology*, 44, 1421–1429. <https://doi.org/10.1037/a0012845>
- Vives, M. L., Repke, L., & Costa, A. (2018). Does bilingualism really affect social flexibility? *Bilingualism: Language and Cognition*, 21, 952–956. <https://doi.org/10.1017/S1366728918000123>
- Ybarra, O., Burnstein, E., Winkielman, P., Keller, M. C., Manis, M., Chan, E., & Rodriguez, J. (2008). Mental exercising through simple socializing: Social interaction promotes general cognitive functioning. *Personality and Social Psychology Bulletin*, 34, 248–259. <https://doi.org/10.1177/0146167207310454>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.