Bilinguals’ inference of emotions in ambiguous speech

Marie-France Champoux-Larsson
Mid Sweden University, Sweden

Alexandra S Dylman
Stockholm University, Sweden

Abstract

Aims and objectives: This study aimed to establish whether adults have a preference for semantics or emotional prosody (EP) when identifying the emotional valence of an utterance, and whether this is affected by bilingualism. Additionally, we wanted to determine whether the prosodic bias (PB) found in bilingual children in a previous study persisted through adulthood.

Design: Sixty-three adults with varying levels of bilingualism identified the emotional valence of words with positive, negative or neutral semantics expressed with a positive, negative, or neutral EP. In Part 1, participants chose whichever cue felt most natural to them (out of semantics or prosody). In Part 2, participants were instructed to identify either the semantics or the prosody in different experimental blocks.

Data and analysis: In Part 1, a one-sample t-test was used to determine whether one type of cue was preferred. Furthermore, a linear regression was used with the participants’ language profile score (measured with the Language and Social Background Questionnaire, LSBQ) as a predictor and how often prosody was chosen as the outcome variable. In Part 2, we ran a linear regression with the LSBQ score as the predictor and a PB score as the outcome.

Findings: In Part 1, participants chose semantics and prosody equally often, and the LSBQ score did not predict a preference for prosody. In Part 2, higher LSBQ scores lead to a larger PB.

Originality: This is the first study to show that bilingual adults, like children, have an increased bias towards EP the more bilingual they are, but only under constrained experimental conditions.

Implications: This study was the first to empirically investigate the conscious choice of emotional cues in speech. Furthermore, we discuss theoretical implications of our results in relation to methodological limitations with experimental settings in bilingual research.

Keywords

Emotional speech, prosody, semantics, bilingualism, prosodic bias

Corresponding author:
Marie-France Champoux-Larsson, Department of Psychology and Social Work, Mid Sweden University, Kunskapens väg 1, Östersund, 831 25, Sweden.
Email: mfclarsson@gmail.com
Introduction

Communication is a complex interplay where we have to process a large amount of information simultaneously. For instance, when it comes to understanding the emotions of an interlocutor with whom we are interacting face-to-face, we have to simultaneously observe that person, listen to the content of their utterances, and how they express what they are saying, all this while it is occurring in a distinct social context with its specific social rules and culture. Research shows that when several of the channels, for instance tone of voice, facial expression, and the actual content of what is said, are present and congruent, interpretation is much easier than when only one of those channels are present (Paulmann & Pell, 2011).

Moreover, when speech is ambiguous, for example when the semantics and prosody of an utterance are incongruent, it is often assumed that people will rely on the prosody to infer the speaker’s intention. For instance, an early study by Mehrabian and Wiener (1967) seems to support the idea that prosody prevails over semantics when the two are incongruent. In the Mehrabian and Wiener study, participants had to determine the feelings of a speaker towards an addressee based on utterances with incongruent semantics and prosody by either paying attention to both semantics and prosody, to semantics only, or to prosody only. Mehrabian and Wiener concluded that prosody trumps semantics, although a closer look at their results shows that semantics also was statistically significant in several conditions. Furthermore, due to the complexity of the design along with a small sample (i.e., a $3 \times 3 \times 2$ between-subject design with 10 participants in each group), their results should be interpreted with caution. However, in a more recent study by Morton and Trehub (2001), adult participants tended to rely more on prosody than semantics when asked to interpret the emotional state of a speaker when listening to utterances with incongruent semantics and emotional prosody (EP). Note, however, that in the instructions, participants were asked to listen carefully to the speaker’s voice, which potentially may have been interpreted as being the same thing as tone of voice.

Evidence from research on empathy has found that semantics are central for empathic accuracy when participants are asked to more generally determine an interlocutor’s emotions. For instance, Hall and Mast (2007) found that when a listener has to interpret an interlocutor’s thoughts and feelings, the most useful information is what is actually said (i.e., semantics), followed by EP, and finally by other visual non-verbal cues. Similarly, Gesn and Ickes (1999) found that empathic accuracy was mainly dependant on verbal (especially the actual content of what was being said), rather than non-verbal cues. These results from empathic accuracy research suggest that prosody is not necessarily the most useful cue to use when accurately interpreting an interlocutor’s feelings. Given that some of the studies presented above indicate that listeners rely more on semantics, while other studies presented here suggest that prosody is the go-to cue when interpreting incongruent speech, it is still an open question whether incongruent utterances are interpreted based on semantics, prosody, or even both.

More recent studies, however, have suggested that prosody is the “appropriate cue” to establish the true meaning of a speaker (e.g., Hellbernd & Sammler, 2016; Yow & Markman, 2011). However, such studies either make this fundamental assumption, or investigate occurrences where prosody is the correct answer, such as sarcasm. But even when using sarcasm and irony, prosody may not always be the predominant cue used by the listener to determine the speaker’s intention. For instance, Rivière et al. (2018) found that a majority of their participants relied mainly on contextual information that was provided before an ironic or non-ironic utterance was presented in order to determine whether an utterance was ironic or not. More specifically, when the context suggested a non-ironic situation, participants judged the utterance following the presentation of the contextual information as non-ironic, even if the utterance was expressed with an
ironic tone of voice. Thus, although prosody is clearly an important cue to help decipher a speaker’s intended meaning, in some ambiguous or incongruent situations, prosody may not be as important as many have assumed.

Indeed, Misono et al. (1997) found that when listening to ambiguous sentences, the preference for semantic cues or prosody varies depending on the context. In their study, the authors presented utterances where, grammatically speaking, there were two possible subjects for an action or two possible owners of an object. They found that when prosody stressed a second but less likely subject or owner in a sentence (e.g., prosody stressing “the lady” in the utterance “The perpetrator threatened the lady with the knife”), nearly 77% of the participants still interpreted the first (and more likely) subject as the intended subject. Hence, a large majority of the participants based their judgement of the ambiguous sentence on its semantic information despite a discrepant prosody which, according to some, should indicate and be interpreted as the true intended meaning. Misono et al. (1997) clearly show that it is not always the case, and that other factors can reduce the influence of prosody on interpretation.

However, it is important to note that there are important cultural differences between different types of languages. For instance, in most European languages, there is a tacit assumption that what is said is also what is meant, as opposed to several Asian languages such as Japanese where what is meant depends on how it is said (Ishii & Kitayama, 2002). Unsurprisingly, research shows that listeners of Japanese rely primarily on prosody when identifying their interlocutor’s meaning when utterances have an incongruent content and tone of voice (Ishii & Kitayama, 2002). The same was found in a population of Mandarin Chinese speakers where prosody was always more salient than semantics in an auditory emotional Stroop task (Lin et al., 2020). Thus, both the context of the utterance (as in Misono et al., 1997) and the cultural context make the interplay between different cues during communication quite complex.

Other factors could also influence which type of cue is used to infer intended meaning in speech. For instance, Yow and Markman (2011) suggested that bilingual children are better at paying attention to non-verbal emotional verbal cues due to their increased need (compared to monolinguals) to pay attention to their surroundings and interlocutor in order to choose the appropriate language to interact in. Indeed, they found that preschoolers that were bilingual were better at interpreting emotions in speech based on prosody compared to monolingual children, who were more inclined to use semantics. However, an underlying assumption in Yow and Markman (2011), that was based off Morton and Trehub’s (2001) findings, was that EP is the cue that reveals a speaker’s feelings more than semantics. Yet, based on the research presented above, this is not robustly established and appears to vary greatly depending on the context. Thus, what was shown by Yow and Markman was that bilingual children differed in their use of emotional cues in speech compared to monolingual children.

Importantly, a recent study by Champoux-Larsson and Dylman (2019) suggests that one’s level of bilingualism affects which type of cue (semantics or prosody) attention is directed towards when emotional words are uttered with a discrepant EP (such as a word with a positive semantic meaning uttered in an angry tone of voice), at least in children. In their study, the authors asked children with varying levels of bilingualism (bilingualism was measured on a continuous scale from monolingual to bilingual) to determine the valence of spoken words based specifically on either their semantics or on their EP. The crucial finding was that, while children performed similarly overall, the more bilingual children were, the more difficulty they had ignoring the distractor on trials where semantics was the target and prosody was the distractor (i.e., they made more mistakes based on the distractor). In contrast to Yow and Markman (2011), Champoux-Larsson and Dylman (2019) did not assume that prosody is always the correct answer and did not interpret their results as an advantage in bilingual children, but rather as a bias towards EP in relation to the degree of
bilingualism. Although it is not clear which mechanisms are behind these results, this study, similar to Yow and Markman’s (2011) study, suggests that bilinguals process cues in emotional speech differently as compared to monolinguals. Yet, it is not clear whether bilingualism leads to an advantage in the processing of EP, a preference, or a bias, particularly in adults.

Thus, two important questions arise from the study by Champoux-Larsson and Dylman (2019). Firstly, since all the participants were children, it is unclear whether this proposed bias towards prosody persists through adulthood. Indeed, there are several examples in the bilingual literature where differences between monolinguals and bilinguals are found during childhood only to disappear in adulthood (see Bialystok et al. (2005) for an example on inhibitory control throughout the lifespan). Extrapolating an effect found in bilingual children to bilingual adults can therefore not be done without first being empirically investigated. For instance, Bhatara et al. (2016) found that the more proficient participants were in their second language (L2), the less accurate they were at identifying positive emotions based on EP in neutral utterances in their L2. However, in Bhatara et al. (2016), only the prosody of utterances provided emotional information, while semantics were held neutral. Thus, while the authors could investigate the accuracy with which bilingual participants interpreted EP in their L2, the design did not allow to investigate potential preferences for either semantics or prosody (Bhatara et al., 2016). Secondly, there is an additional, and perhaps more crucial, question that arises from Champoux-Larsson and Dylman’s (2019) study. Namely, those participants were specifically asked to base their responses on either the semantics or the prosody of the words. Thus, it is unclear whether participants would base their responses on semantics or prosody in a free choice task (which might reflect a level of ambiguity more similar to real life), and whether this preference would depend on their level of bilingualism. In other words, it is unclear which cue a bilingual would choose to interpret emotion in speech if they were not instructed to rely on either semantics or on prosody, as was done in Yow and Markman (2011).

In light of the abovementioned studies, the current study investigated which type of cue (semantics or prosody) adult participants base their judgement on when listening to words with a semantic and prosodic emotional content in particular, and whether bilingualism affects this choice. Also, we investigated whether the prosodic bias (PB) found in bilingual children also exists in adult bilinguals. To investigate these two areas, we first asked adults with varying levels of bilingualism (from mostly monolingual to mostly bilingual) to determine the emotional valence of utterances based on their general impression (i.e., without specifying which cue to use) to determine if there is a preferred cue and whether it is moderated by bilingualism. In Part 2, we asked them to determine the utterance’s valence based on its EP or on its semantic content specifically to determine whether a PB also exists in adult bilinguals.

**Methods**

**Participants**

Participants were recruited through advertisements on the department’s social media and directly on campus. The total sample consisted of 74 participants (25.7% males, 73% females, 1.3% other) aged 18 to 50 years (mean ($M$) = 25.93, standard deviation ($SD$) = 7.29). However, one participant reported not having normal or corrected hearing, and 10 participants reported having started to learn Swedish after age five (thus making Swedish an L2 to them), and were excluded from further analysis. The final sample consisted of 63 native Swedish speakers (28.6% males, 69.8% females, 1.6% other) aged 18 to 50 years ($M = 25.95$, $SD = 7.17$) reporting having English as their L2, or one of their L2s if they had more than one.
The participants’ language profile was measured using Parts B and C of the Language and Social Background Questionnaire (LSBQ: Anderson et al., 2018). Specifically, the Composite Factor Score (CFS), which is the score developed by Anderson et al. (2018), was computed via the provided score calculator. The CFS is a complete score that includes a large array of important facets of bilingualism, namely, proficiency in the respondent’s first language (L1) and L2, frequency of use of the L1 and L2, which language(s) was heard during different periods in life (from infancy to adolescence) and from different people (parents, siblings, grand-parents, relatives, partner, roommates, neighbours, and friends), which language(s) is used in different contexts (at home, school, work, for social activities, religious activities, hobbies, shopping, and social services), and for different activities (reading, writing emails or text messages, in social media, to write lists, when watching TV or movies, listening to the radio, surfing on the internet, and praying), as well as frequency of code-switching (with family, friends, and on social media). The CFS can be used as a continuous variable, or can be split into distinct groups based on the recommendations found in the LSBQ’s score calculator. Here, the CFS was used as a continuous variable since it is increasingly argued that this way to operationalise bilingualism better reflects the true nature of the concept (e.g., DeLuca et al., 2019; Edwards, 2012; Gullifer et al., 2018; Gullifer & Titone, 2020; Incera & McLennan, 2018; Jylkkä et al., 2017; Kaushanskyaya & Prior, 2015; Luk & Bialystok, 2013; Sulpizio et al., 2020; Surraion & Luk, 2019), and because we wanted to replicate Champoux-Larsson and Dylman (2019) as closely as possible. In this study, level of bilingualism thus refers to the computed CFS for each participant based on their answers on the LSBQ. In other words, our sample consisted of participants with varying levels of bilingualism, with participants being more monolingual at the lower end of the scale, and participants being more bilingual at the upper end of the scale ($M = 4.74, SD = 3.63$). We controlled with bivariate correlation analyses that age and highest level of completed education (on a scale from 1 to 6 where 1 = elementary school or lower, 2 = high school, 3 = professional education, 4 = Bachelor’s degree, 5 = Master’s degree, 6 = PhD: $M = 2.49, SD = 0.98$) respectively did not correlate with the CFS. Both analyses were non-significant (age: $r = -0.001, p = 0.992$; education: $r = 0.092, p = 0.475$). All questions were presented and responded to via the survey platform Qualtrics.

**Stimuli**

The stimuli from Champoux-Larsson and Dylman (2019) were used in the current study. These consisted of 108 different recordings of 18 single words with a positive (e.g., love), negative (e.g., dead), or neutral (e.g., clock) semantics (six words per valence) in Swedish. All words were presented vocally and uttered in a positive (happy), negative (angry) and neutral tone of voice by one female and one male native speaker, resulting in the 108 recordings. Of the 108 recordings, the valence of the semantics and prosody were congruent for 36 of them, and the valence of the semantics and prosody were incongruent for the remaining 72 recordings. As reported in Champoux-Larsson and Dylman (2019), the words were selected from lists of words that had previously been produced in a pilot study based on different emotional categories (positive, negative, and neutral) and were afterwards rated by other independent raters based on valence, arousal and dominance. The authors controlled that words were matched on arousal, frequency, and number of letters. Furthermore, the recordings created by Champoux-Larsson and Dylman were validated by independent raters until an inter-rater agreement of at least 0.8 was reached for the EP (see detailed information on the selection and validation of the stimuli in Champoux-Larsson and Dylman (2019)).
Design

There were two parts in this experiment: non-directed; and directed. In the non-directed part (Part 1), participants grounded their judgement of the valence of the stimulus based on what felt most natural to them (semantics or prosody). The 108 recordings were presented randomly with two breaks, one after 36 trials and the other after 72 trials. Between each trial (i.e., a recording being played for the participant followed by the participant’s response), a fixation cross was presented for 500 milliseconds (ms). Part 1 was always presented first in order to avoid a priming effect from the second part (which was directed).

In the second part of the experiment, four blocks with directed trials were presented. Participants were asked to base their judgements on either semantics or prosody and to ignore the irrelevant cue (i.e., prosody in the semantics blocks, and semantics in the prosody blocks). The same 108 recordings as in Part 1 were presented (a trial here again consisted of a recording played for the participant, followed by the participant’s response, and a fixation cross for 500 ms before the following trial), and divided into four semi-randomised blocks with valence of word content, valence of tone of voice, congruence between word content and tone of voice, and gender of speaker balanced across the blocks. Two of the blocks were directed towards semantics (i.e., participants had to identify the valence of the words based on their semantics while ignoring the prosody) and two of the blocks were directed towards prosody (i.e., participants had to identify the valence of the utterances based on their prosody while ignoring the semantics). The four blocks were presented in a counterbalanced order thus creating four versions of the experiment to which participants were randomly assigned. SuperLab (version 5) was used for programming and running the experiment on a MacBook Air.

Procedure

Participants were met in the laboratory on campus. They first filled out the survey with background questions and the LSBQ (Anderson et al., 2018). Afterwards, participants completed the computerised task where instructions were provided in writing. Before the non-directed block, there were six practice trials with different recordings that were not used subsequently in the experimental trials. Afterwards, the 108 experimental trials were presented in a randomised order. In order to avoid priming the participants or teaching them the “correct” answer, no feedback on their answers was provided since the aim was to investigate what felt most natural to them. Responses were provided by pressing a drawing depicting a happy (for positive), neutral or angry (for negative) face on the keyboard (placement of the happy and the angry drawings was counterbalanced across participants). For the non-directed block, participants received the following written instructions (in Swedish): “You will listen to words. Indicate whether you interpret each word that you hear as positive, neutral, or negative by pressing the corresponding icon on the keyboard. Answer as quickly and as accurately as possible”. The order of the answer alternatives in the instructions, namely negative, neutral, or positive, corresponded to the order of the icons on the keyboard. After the non-directed block, participants continued with the directed blocks. Before each block, written instructions were provided and the same practice trials as in Part 1 were presented. Answers were provided in the same manner as in the non-directed block, namely by pressing a drawing depicting a happy (for positive), neutral or angry (for negative) face on the keyboard (placement of the happy and the angry drawings was the same as in the non-directed block). The instructions (also in Swedish and in writing) for the semantics blocks were as follows: “This time, indicate whether the MEANING of each word is positive, neutral, or negative by pressing the corresponding icon on the keyboard. Answer as quickly and accurately as possible”. For the prosody blocks, the instructions
were the same except for the beginning, which read: “This time, indicate whether each word is EXPRESSED in a positive, neutral, or negative manner by pressing the corresponding icon on the keyboard”. There was no time limit to provide an answer throughout the experiment. Participants received either course credits or a movie ticket for their participation.

Results

Part 1: Non-directed block

We first conducted a one-sample t-test to ensure that our sample performed above chance (i.e., 55.6% since there were three possible answers of which two were correct for the incongruent trials) on all trials. Participants responded significantly above chance ($M = 83.83$, $SD = 10.71$), $t(62) = 17.65, p < 0.001, d = 2.22$. Subsequently, only incongruent trials were of interest in the non-directed block since congruent trials did not allow to differentiate which cue a participant had based their judgement on (i.e., for a stimulus with positive semantics and positive prosody, it was impossible to know if a participant answered “positive” based on the semantics, prosody, or both). On the other hand, incongruent trials allowed us to make this differentiation. For incongruent trails, since both semantics and prosody were valid choices, there were two potential correct answers, and only one incorrect answer. For instance, a semantically negative word uttered with a neutral tone of voice was judged correctly if the participant answered either negative or neutral since they were free to choose whatever cue they wanted. However, if a participant answered “positive” for a semantically negative stimulus with a neutral prosody, this was considered a mistake. Because we wanted to examine which cue participants primarily based their judgement on, only incongruent trials where participants had provided a correct answer (namely, an answer where the participant provided a correct answer based on either the utterance’s semantics or its prosody, since both alternatives were valid) were included in the analysis. Answers that were faster than 200 ms were considered as errors and were excluded. To account for the different number of trials computed per participant due to the varying number of excluded trials, frequencies were converted into percentages of correct answers for prosody-based correct answers.

We first analysed whether prosody was chosen more often than semantics. If no preference exists, both cues should be chosen approximately 50% of the time. A one-sample t-test revealed that the percentage of correct answers where prosody was chosen ($M = 56.75$, $SD = 33.7$) was not significantly different from 50%, $t(62) = 1.6, p = 0.117$. Next, we used the CFS as a predictor in a linear regression analysis where the percentage of correct answers where prosody was chosen was used as the outcome variable. If increased bilingualism leads to a preference for prosodic cues, the percentage of times when this cue was chosen should increase along with the CFS. The model was not significant, $F(1, 61) = 2.37, p = 0.129, R^2 = 0.037$, suggesting that the degree of bilingualism does not predict a heightened preference for prosodic cues.

Part 2: Directed blocks

Again, we first controlled that our participants had performed above chance levels (i.e., 33% since there were three different possible answers and that only one of them was correct for any given trial). A one-sample t-test showed that they did ($M = 75.83$, $SD = 15.44$), $t(62) = 20.47, p < 0.001, d = 2.58$. To investigate whether a PB exists in adulthood, a linear regression was performed with the mistakes in the semantics blocks (i.e., the block where the participants were asked to report the valence of the word content) that were biased towards the distractor (i.e., the utterance’s prosody) as the outcome variable (in percentage to account for the different number of trials that were
included for each participant). Again, the CFS was the predictor. The model was significant and revealed that the higher the CFS was, the more participants tended to make mistakes that were biased towards the prosody of the utterances in the semantics block, \( F(1, 61) = 4.37, p = 0.041, R^2 = 0.07 \).

As the PB may reflect a difficulty in ignoring distractors in general, we also controlled whether the opposite bias, namely a bias towards semantics (when the target was the prosody and the distractor was the semantics) existed. A semantic bias was computed by calculating the percentage of mistakes that were biased towards the distractor (semantics) in the prosody blocks. A linear regression analysis with CFS as predictor was not significant, \( F(1, 61) = 0.28, p = 0.596, R^2 = 0.005 \).

In order to control for differences in general performance, as the PB may reflect a poorer general performance the more bilingual the participants were, the participants’ accuracy on the task in general was investigated. A linear regression with accuracy for all conditions as the outcome variable and the CFS as the predictor was not significant, \( F(1, 61) = 1.61, p = 0.21, R^2 = 0.03 \), suggesting that the participants’ level of bilingualism did not affect their performance in general.

Furthermore, the bilinguals’ tendency to attend to prosody more has been interpreted as a bilingual advantage in prosody processing in Yow and Markman (2011), while Champoux-Larsson and Dylman (2019) posit that this effect may in fact be caused by the bias towards prosody that they found. In order to verify whether our sample performed better on incongruent trials where prosody was the target, which would reflect an advantage in prosody processing, the percentage of correct responses on incongruent trials in the prosody blocks was analysed with a linear regression analysis using the CFS as predictor. The model was not significant, \( F(1, 61) = 0.001, p = 0.98, R^2 < 0.001 \), suggesting that bilingualism does not lead to an advantage in performance when identifying EP in incongruent utterances.

Finally, for exploratory purposes only, we investigated general performance in terms of accuracy on congruent semantics and congruent prosody trials, as well as on incongruent semantics and incongruent prosody trials. A paired-sample \( t \)-test revealed that participants performed better on congruent semantics trials (\( M = 14.13, SD = 3.19 \)) than on congruent prosody trials (\( M = 13.11, SD = 2.87 \)), \( t(62) = 2.55, p = 0.013, d = 0.32 \). As for incongruent trials, a paired-sample \( t \)-test showed that participants again were more accurate for semantics (\( M = 26.44, SD = 7.97 \)) than for prosody trials (\( M = 22.14, SD = 7.32 \)), \( t(62) = 3.41, p = 0.001, d = 0.43 \). We also explored the mean reaction times by conducting a two-way repeated measures analysis of variance with type of cue (semantics, prosody) and congruence (congruent, incongruent) as independent variables, and reaction times as dependant variable. The reaction times did not differ significantly between semantics (\( M = 738, SD = 367 \)) and prosody trials (\( M = 746, SD = 360 \), \( F < 1 \). However, the main effect of congruence was significant here as well, with congruent trials (\( M = 687, SD = 323 \)) being responded to faster than incongruent trials (\( M = 796, SD = 392 \)), \( F(1, 72) = 17.65, p < 0.001, \eta^2 = 0.069 \). The interaction between type of cue and congruence approached significance, but was not significant, \( F(1, 72) = 3.5, p = 0.065, \eta^2 = 0.012 \). All in all, these analyses suggest that semantics is easier to interpret than prosody in terms of accuracy, albeit not slower, and that incongruent trials are more effortful than congruent trials.

**Discussion**

This study investigated the type of cue (semantics or prosody) that adults tend to rely on to determine emotional valence when listening to words that are positive, negative or neutral and uttered with an incongruent prosody, both in general and as a function of their language profile. It also investigated whether the proposed PB found in children in Champoux-Larsson and Dylman (2019) persists into adulthood. The results suggest that, in general, semantics and prosody are chosen
equally often to determine an utterance’s emotional valence when the semantics and EP are incongruent, and that it is not modulated by the extent to which a person is bilingual. Furthermore, our results show that the PB found during childhood in Champoux-Larsson and Dylman (2019) is also found in adult bilinguals. Namely, the more bilingual participants were (based on the CFS), the more they tended to make mistakes biased towards the prosody of utterances in constrained task settings when the correct answer should have been the semantics. Taken together, our results suggest that prosody is processed or paid attention to differently the more bilingual a person is, but only under constrained conditions. From a developmental point of view, this study also suggests that bilinguals follow a particular developmental path regarding the processing of emotion in speech, at least when it comes to prosodic cues. Indeed, the effect found in children in Champoux-Larsson and Dylman (2019) was virtually the same PB that we found in this study in an adult population. A main difference however is that while the children in Champoux-Larsson and Dylman performed better on prosody trials the more bilingual they were, this was not the case with our adult population. Note however that we cannot establish this effect without doubt since the two groups belonged to two different studies and could therefore not be compared directly, since this study was not longitudinal.

The pattern of results from the current study has parallels with other phenomena in bilingual research. For instance, the code-switching literature has repeatedly shown a “mixing cost” when a speaker is instructed to switch from one language to another, a process which is effortful, recruits more resources and leads to longer reaction times (e.g., Jevtović et al., 2019; Kleinman & Gollan, 2016). This may seem counterintuitive since bilinguals otherwise appear to switch effortlessly and seamlessly from one language to another in settings where they are free to speak whatever language they want. Indeed, code-switching has been suggested to reflect a process which allows communication to be less costly and more efficient (e.g., Kleinman & Gollan, 2016). In fact, recent studies show that when a person is obliged to switch during an experimental task, a mixing cost clearly appears, but that when switching is voluntary during the same task, no such cost emerges (e.g., Blanco-Elorrieta & Pylkkänen, 2017; de Bruin et al., 2018; Jevtović et al., 2019). Together with the findings of the current study, these studies indicate that the effects that are observed in less natural and constrained settings do not necessarily reflect the reality of bilingual communication in real life. This is important as much knowledge is built on effects that are found in controlled and constrained contexts. In the current case, had we only used a constrained and directed condition (Part 2 of the study), our results would have suggested a fundamental difference between monolinguals and bilinguals in emotion processing in speech. However, because of the non-directed condition (Part 1 of the study), more nuanced conclusions can be drawn, suggesting that although increased bilingualism leads to differences in the processing of EP in a constrained setting, it may not necessarily have substantial consequences in real life situations.

Although our study cannot establish the underpinnings of this effect, given its relationship with the level of bilingualism that our participants had, one could tentatively speculate about the interference that the participants’ L2, particularly English, had on their performance. However, the current data cannot give any clear indications regarding this for several reasons. Firstly, while all participants had English as an L2 (due to English being a compulsory subject in the Swedish school), the participants’ English language proficiency likely varied greatly across the sample, especially given that we measured bilingualism on a continuous scale. As the entire experiment was conducted in Swedish only (which was the participants’ native language), we did not measure their English language proficiency specifically. While the LSBQ does technically ask about L2 proficiency, this is only measured using four sub-questions, and does not take into consideration multiple L2s.
Secondly, several of the participants reported additional L2s apart from English, even if all of them reported having at least English as an L2. It is, therefore, difficult to comment on the level of influence from the participants’ L2s in a reliable way. Even if we did take L2 proficiency into consideration, given that the reported L2s themselves varied greatly across participants, where some of the L2s were from the same language families whereas others were not, measuring the level of emotional influence from the L2 in this context would be speculative at best. Despite the variability in L2s, however, we did find an effect of bilingualism in Part 2, which strengthens the generalisability of this study to various types of bilinguals. However, a more stringent sample with a specific and homogeneous language profile could potentially affect the results, particularly in Part 1 of our study.

Furthermore, while there are studies showing parallel activation of bilinguals’ two languages, these have mainly shown cross-linguistic influence of the native language on the L2 (e.g., Thierry & Wu, 2007). In contrast, the participants in our study completed the task in a strict L1 context. Studies specifically investigating cross-linguistic influence in both L1 and L2 have found an asymmetrical pattern of results whereby the L1 influences L2 naming while the influence from L2 on L1 naming is considerably smaller (e.g., Dylman & Barry, 2018). Of course, these studies have not investigated emotion words per se, and more recent research on, for example, the foreign language effect in decision-making, has indicated a potential transfer of emotional resonance from linguistically similar languages such as Swedish and Norwegian (e.g., Dylman & Champoux-Larsson, 2020), and so, these issues may need to be more closely investigated in future studies.

Another important issue is the consequences of investigating and interpreting results in different ways to support or refute the debated bilingual advantage concept. As Champoux-Larsson and Dylman (2019) showed in their study, what was originally interpreted as a bilingual advantage in the processing of EP was driven by a bias towards prosody. In the current study, we did not find that higher bilingualism scores led to better performance on the prosody trials (i.e., we did not find a bilingual advantage), but we still replicated the PB. If the PB had not been investigated in Champoux-Larsson and Dylman (2019) or in this study, only the development of the alleged bilingual advantage would have been the focus of both studies. Namely, Champoux-Larsson and Dylman would likely have claimed to have replicated the bilingual advantage in prosody processing in children, and we would simply have concluded that the advantage in prosody processing found in bilingual children disappears in adulthood (similarly to what other studies investigating the development of alleged bilingual advantages in other domains usually find, see for example Bialystok et al., 2012). However, because the study by Champoux-Larsson and Dylman (2019) and this study also analysed the types of mistakes that the participants make, both studies show that the reality of prosody processing, in constrained contexts, is more intricate and complex than what a bilingual advantage approach could explain on its own.

Furthermore, as in Yow and Markman (2011), participants were not instructed to focus specifically on one of the two cues in Part 1 of this study. Unlike in Yow and Markman however, we asked our participants in Part 2 to focus specifically on one of the two cues, thus creating a distinct distractor. On the other hand, unlike Yow and Markman, where they coded the answers as correct when they were based on paralanguage derived from the results in Morton and Trehub (2001), we did not assume that prosody was the correct answer in Part 2. Nevertheless, even if we had done so, the percentage of correct answers based on prosody in Part 1 was not modulated by bilingualism. Taken together, the lack of effect of bilingualism on performance for prosody processing in Parts 1 and 2 suggest that no so-called bilingual advantage could be replicated. On the other hand, it supports the idea that bilingualism leads to a bias towards prosody in adults just as it does in children. Simultaneously, our results do not necessarily mean that bilingualism leads to a better processing of prosody during childhood. The fact that no effect on accuracy for prosody processing
was found in this study compared to Yow and Markman (2011) and Champoux-Larsson and Dylman (2019), where the participants were children, could simply be a reflection of normal cognitive development given that the participants were adults in the current study. Note, however, that the semantics and prosody blocks in Part 2 were presented in alternation. Since we did not subsequently remind participants which cue to focus on after the instructions were given at the beginning of each block, we cannot rule out completely that some participants were confused and forgot about the instructions. However, since the blocks started with clear instructions and practice trials, and because they were relatively short in terms of duration, it is improbable that participants had time to forget about the instructions. Thus, even though we cannot rule out this possibility completely, this limitation in our design is unlikely to have affected our results.

Additionally, the effect size of the PB that we found was quite small, and results should therefore be interpreted cautiously, particularly when it comes to real life effects considering the artificial setting that we tested our participants in. The fact that our results in Part 1 found that semantics and prosody are chosen equally often, and the fact that our analyses in Part 2 only found a small effect of bilingualism with the PB indicates that all cues are important when interpreting the meaning of what is said by an interlocutor. Also, the results that were found using a laboratory setting are likely to differ in a real life setting where other aspects such as attention, motivation, and context will affect how and with what cues a person interprets a talker’s intention and emotions. All in all, however, our study suggests that there may be different mechanisms underlying the processing of EP in speech as one is increasingly bilingual, at least to some extent and in some contexts. However, our design does not allow explaining which mechanisms are involved. A possibility is that, as Champoux-Larsson and Dylman (2019) hypothesise, prosody across languages may show less variability than semantics, thus making prosody more constant, more reliable, and thus less effortful to use, but only for the most bilingual individuals. This could be a consequence of the different social demands that bilinguals face. While monolinguals only have one language to choose from, bilinguals must constantly monitor their interlocutor in order to determine which language(s) to use. These different social monitoring demands may lead to more permanent differences in how prosody is processed. However, more research will be needed to understand the underpinnings of the effects found in this study.

Finally, while we did not specifically examine or measure executive functioning, executive functions are likely involved, particularly in Part 2, where the participants were task bound to specifically attend to one cue while actively ignoring a distractor. This task naturally must involve some degree of cognitive control. There is a vast literature on executive functioning in bilinguals (see, e.g., Costa et al., 2009; Green & Abutalebi, 2013; Luk et al., 2012; Soveri et al., 2011), as well as an extensive literature and debate regarding the existence of the so-called bilingual advantage in executive functioning (e.g., Bialystok, 2011; Bialystok et al., 2012; Hilchey & Klein, 2011; Lehtonen et al., 2018; Paap & Greenberg, 2013; Paap et al., 2014). However, there are many methodological issues raised in this debate. For example, these studies have predominantly investigated (different populations of) balanced, or simultaneous, bilinguals, whereas the current study purposefully examined a more heterogeneous sample of bilinguals measuring bilingualism on a continuous scale. Indeed, a recent study has shown that different operationalisations of bilingualism can, in fact, yield different results in the same sample of bilinguals conducting an executive function task (Champoux-Larsson & Dylman, 2021). Thus, much remains to be examined with regard to the role of executive functions (including which one or which ones) in the bilingual literature, how best to define and operationalise bilingualism in future studies, and even how we can go about designing tasks that actually measure relevant executive functions in the first place. These are important methodological and theoretical questions going forward. Likewise, an interesting future direction from this study is to specifically investigate how executive functions are involved in
determining the emotional state of an interlocutor based on semantics and prosody, and perhaps even more importantly, the interaction between semantics, EP, but also context.

Acknowledgements
The authors thank Jennifer Hjort and Ekaterina Wickström for their assistance with collecting and preparing data for analysis.

Declaration of conflicting interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Marie-France Champoux-Larsson https://orcid.org/0000-0001-7251-5263

References
Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. Behavior Research Methods, 50(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Bhatara, A., Laukka, P., Boll-Avetisyan, N., Granjon, L., Elfenbein, H. A., & Bänziger, T. (2016). Second language ability and emotional prosody perception. PLoS One, 11(6), Article e0156855. https://doi.org/10.1371/journal.pone.0156855

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. Journal of Experimental Child Psychology, 110(3), 461–468. http://doi.org/10.1016/j.jecp.2011.05.005

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. Trends in Cognitive Science, 16(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. International Journal of Bilingualism, 9(1), 103–119. https://doi.org/10.1177/13670069050090010701

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. Journal of Neuroscience, 37, 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017

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

Champoux-Larsson, M.-F., & Dylman, A. S. (2019). A prosodic bias, not an advantage, in bilinguals’ interpretation of emotional prosody. Bilingualism: Language and Cognition, 22(2), 416–424. https://doi.org/10.1017/S1366728918000640

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

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why to bilinguals switch between their languages? Journal of Memory and Language, 103, 28–43. https://doi.org/10.1016/j.jml.2018.07.005

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. Proceedings of the National
Academy of Sciences of the United States of America, 116(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translations in bilinguals and synonyms in monolinguals. Cognition, 171, 151–171. https://doi.org/10.1016/j.cognition.2017.09.014

Dylman, A. S., & Champoux-Larsson, M.-F. (2020). It’s (not) all Greek to me: Boundaries of the foreign language effect. Cognition, 104148. https://doi.org/10.1016/j.cognition.2019.104148

Edwards, J. (2012). Conceptual and methodological issues in bilingualism and multilingualism research. In T. K. Bhatia, & W. C. Ritchie (Eds.), The handbook of bilingualism and multilingualism (pp. 5–25). Wiley-Blackwell.

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. Journal of Personality and Social Psychology, 77(4), 746–761. https://doi.org/10.1037/0022-3514.77.4.746

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. Journal of Cognitive Psychology, 25(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. Neuropsychologia, 117, 123–134. https://doi.org/10.1016/j.neuropsychologia.2018.04.037

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. Bilingualism: Language and Cognition, 23(2), 283–294. https://doi.org/10.1017/S1366728919000026

Hall, J. A., & Mast, M. S. (2007). Sources of accuracy in the empathic accuracy paradigm. Emotion, 7(2), 438–486. https://doi.org/10.1037/1528-3542.7.2.438

Hellbernd, N., & Sammller, D. (2016). Prosody conveys speaker’s intentions: Acoustic cues for speech act perception. Journal of Memory and Language, 88, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. Psychonomic Bulletin & Review, 18(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. Aging, Neuropsychology, and Cognition, 25(3), 443–463. https://doi.org/10.1080/13825585.2017.1319902

Ishii, K., & Kitayama, S. (2002). Processing of emotional utterances: Is vocal tone really more significant than verbal content in Japanese? Cognitive Studies, 9(1), 67–76. https://doi.org/10.11225/jcss.9.67

Jevtović, M., Duñabeitia, J. A., & de Bruin, A. (2019). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. Bilingualism: Language and Cognition, 23(2), 401–413. https://doi.org/10.1017/S1366728919000191

Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodriguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals: An Internet-based study. Journal of Cognitive Psychology, 29(4), 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(1), 27–28. https://doi.org/10.1017/S1366728914000510

Kleinman, D., & Golla, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. Psychological Science, 27(5), 700–714. https://doi.org/10.1177/0956797616634633

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(4), 394–425. https://doi.org/10.1037/bul0000142

Lin, Y., Ding, H., & Zhang, Y. (2020). Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal Stroop effects. Journal of Speech, Language, and Hearing Research, 63(3), 896–912. https://doi.org/10.1044/2020_JSLHR-19-00258
Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes, 27*(10), 1479–1488. https://doi.org/10.1080/01690965.2011.613209

Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology, 6*(1), 109–114. https://doi.org/10.1037/h0024532

Misono, Y., Mazuka, R., Kondo, T., & Kiritani, S. (1997). Effects and limitations of prosodic and semantic biases on syntactic disambiguation. *Journal of Psycholinguistic Research, 26*(2), 229–245.

Morton, J. B., & Trehub, S. E. (2001). Children’s understanding of emotion in speech. *Child Development, 72*(3), 834–843. https://doi.org/10.1111/1467-8624.00318

Paap, K., Johnson, H., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*(6), 615–639. http://doi.org/10.1080/20445911.2014.944914

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232–258. http://doi.org/10.1016/j.cogpsych.2012.12.002

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion, 35*(2), 192–201. http://doi.org/10.1007/s11031-011-9206-0

Rivièrè, E., Klein, M., & Champagne-Lavau, M. (2018). Using context and prosody in irony understanding: Variability amongst individuals. *Journal of Pragmatics, 138*, 165–172. https://doi.org/10.1016/j.pragma.2018.10.006

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Frontiers in Psychology, 2*, Article 183. https://doi.org/10.3389/fpsyg.2011.00183

Sulpizio, S., Del Maschio, N., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2020). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *NeuroImage, 205*, Article 116306. https://doi.org/10.1016/j.neuroimage.2019.116306

Surraim, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition, 22*(2), 401–415. https://doi.org/10.1017/S1366728917000682

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Yow, W. G., & Markman, E. M. (2011). Bilingualism and children’s use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition, 14*(4), 562–569. https://doi.org/10.1017/S136672891000404

Author biographies

Marie-France Champoux-Larsson received her PhD in Psychology at Mid Sweden University, Sweden, in 2018. For her doctoral thesis, she investigated the perception of emotions in speech and faces in bilingual children and adults. She is currently continuing her research as a postdoctoral researcher at Mid Sweden University investigating how language, emotion, and bilingualism interact in social contexts.

Alexandra S Dylman received her PhD in Psychology at the University of Essex, UK, in 2013, looking at language production in bilinguals and monolinguals. She is now working as an Associate Professor at Stockholm University and Mälardalen University in Sweden. Her research area, generally speaking, is within cognition and psycholinguistics, and her primary research interests include bilingualism, language production, reading, and the interaction between language, emotion, and culture.