Bilingualism: Language and Cognition

cambridge.org/bil

Hearing emotion in two languages: A pupillometry study of Cantonese–Mandarin bilinguals' perception of affective cognates in L1 and L2

Yao Yao[1,2], Katrina Connell[1,3,4] and Stephen Politzer-Ahles[1]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, HKSAR, China; [2]The Hong Kong Polytechnic University - Peking University Research Centre on Chinese Linguistics, HKSAR, China; [3]Department of Spanish, Italian, and Portuguese, The Pennsylvania State University, State College, US and [4]Center for Language Sciences, The Pennsylvania State University, State College, US

Research Article

Cite this article: Yao Y, Connell K, Politzer-Ahles S (2023). Hearing emotion in two languages: A pupillometry study of Cantonese–Mandarin bilinguals' perception of affective cognates in L1 and L2. *Bilingualism: Language and Cognition* **26**, 795–808. https://doi.org/10.1017/S1366728922000931

Received: 12 August 2021
Revised: 5 December 2022
Accepted: 12 December 2022
First published online: 17 January 2023

Keywords:
Cantonese–Mandarin bilingual; differential affective processing; pupillometry; lexical processing

Address for correspondence:
Yao Yao,
Email: ctyaoyao@polyu.edu.hk

Abstract

Differential affective processing has been widely documented for bilinguals: L1 affective words elicit higher levels of arousal and stronger emotionality ratings than L2 affective words (Pavlenko, 2012). In this study, we focus on two closely related Chinese languages, Mandarin and Cantonese, whose affective lexicons are highly overlapping, with shared lexical items that only differ in pronunciation across languages. We recorded L1 Cantonese – L2 Mandarin bilinguals' pupil responses to auditory tokens of Cantonese and Mandarin affective words. Our results showed that Cantonese–Mandarin bilinguals had stronger pupil responses when the affective words were pronounced in Cantonese (L1) than when the same words were pronounced in Mandarin (L2). The effect was most evident in taboo words and among bilinguals with lower L2 proficiency. We discuss the theoretical implications of the findings in the frameworks of exemplar theory and models of the bilingual lexicon.

Introduction

Every language has words that are imbued with emotionality. The family of affective words includes (1) emotion words, which describe specific affective states (e.g., *happy*, *sad*) or processes (e.g., *worry*, *rage*), (2) emotion-laden words (e.g., *cancer*, *gift*), (3) taboos (e.g., *rape*), (4) swear words and expletives (*e.g.,* the *f*-word), and (5) reprimands (e.g., *shame on you*), endearments (e.g., *darling*), and interjections (e.g., *ouch*). Affective words can be characterized by the valence (e.g., positive, neutral, negative) and intensity (e.g., high, low) of the emotionality the word is associated with. The group of emotion words is further classified by the taxonomy of human emotions: albeit a still contentious subject in psychological research (see Ekman, 1992; John, 1988; Oatley & Johnson-Laird, 1987; Plutchik, 1980; Strauss & Allen, 2008; Turner, 2000), most researchers agree on a set of basic human emotions such as HAPPINESS, SADNESS, ANGER, FEAR, and SURPRISE, with more complex emotions viewed as a blend of two or more basic emotions.

The emotionality of affective words is deeply embedded. Compared to words with no emotionality, affective words are known to elicit higher levels of arousal when encountered visually or auditorily. The most direct evidence comes from physiological and neural studies: higher levels of skin conductance responses and pupil dilation, as well as distinct neural patterns, are recorded when participants are exposed to emotionally-charged words and phrases compared to neutral ones (Bowers & Pleydell-Pearce, 2011; Caldwell-Harris, 2014; Chen, Lin, Chen, Lu, & Guo, 2015; Eilola & Havelka, 2011; Harris, Ayçiçegi, & Gleason, 2003; Hsu, Jacobs, & Conrad, 2015; Iacozza, Costa, & Duñabeitia, 2017; Kensinger & Corkin, 2004; Kissler & Strehlow, 2017; Maddock, Garrett, & Buonocore, 2003; Manning & Melchiori, 1974; Maratos, Allan, & Rugg, 2000; Mortier, 2013; Opitz & Degner, 2012; Partala & Surakka, 2003; Recio, Conrad, Hansen, & Jacobs, 2014; Simcox, Pilotti, Mahamane, & Romero, 2012). Affective words are also distinguished from neutral words in cognitive processing, demonstrating better memory and recall performance (Anooshian & Hertel, 1994; Kensinger & Corkin, 2004; Maratos et al., 2000; Talmi & Moscovitch, 2004), faster responses in lexical decision and identification (Bayer, Sommer, & Schacht, 2011; Kazanas & Altarriba, 2015; Recio et al., 2014), stronger interference in emotional Stroop tasks (Winskel, 2013), and evidence of affective priming in lexical processing (Altarriba & Canary, 2004). These effects are widely attributed to the automatic emotional arousal triggered by the presentation of affective words, which may either facilitate or inhibit lexical processing depending on the nature of the processing task (however, see Maratos et al., 2000 and Talmi & Moscovitch, 2004 for a discussion of possible confounding factors).

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

What happens if the language user is a bilingual who has access to two lexicons? A general consensus is that bilinguals respond more strongly to affective words in their first language (L1) than those in their second language (L2), a phenomenon often referred to as differential affective processing (for comprehensive reviews, see Caldwell-Harris, 2014, 2015; Pavlenko, 2012; Robinson & Altarriba, 2018). L1-L2 differences are observed when bilinguals are explicitly asked to rate the emotionality of affective words, yielding more extreme ratings of valence and higher ratings of intensity for L1 affective words than the L2 counterparts (Dewaele, 2004, 2008; Garrido & Prada, 2018; Milanović, 2019). More convincing evidence comes from neural/physiological measures of arousal and behavioral measures of cognitive processing performance, as the emotionality effects discussed above often have a greater magnitude in the L1 than the L2 – if present at all. A widely-cited study by Harris et al. (2003) reported that Turkish–English[1] bilinguals had heightened skin conductance responses to taboo words (e.g., *asshole*) and reprimands (*Shame on you!*) in Turkish (L1) compared to the equivalent expressions in English (L2). Altarriba and colleagues (Altarriba, 2003; Altarriba & Canary, 2004; Kazanas & Altarriba, 2016; Sutton, Altarriba, Gianico, & Basnight-Brown, 2007) conducted a series of cognitive processing experiments (emotional Stroop task, masked lexical decision, affective priming) with Spanish–English bilinguals, and found greater emotionality effects with Spanish (L1) stimuli than with English (L2) stimuli. Similar evidence has been reported for a wide range of bilingual profiles (English–French: Segalowitz, Trofimovich, Gatbonton, & Sokolovskaya, 2008; English–Spanish: Simcox et al., 2012; German–English: Hsu et al., 2015; Finnish–English: Eilola, Havelka, & Sharma, 2007; Portuguese–English: Garrido & Prada, 2018; Chinese–English: Caldwell-Harris, Tong, Lung, & Poo, 2011; Fan et al., 2016; Thai–English: Winskel, 2013; Catalan–Spanish: Ferré, García, Fraga, Sánchez-Casas, & Molero, 2010; Ferré, Sánchez-Casas, & Fraga, 2013; German–French: Degner, Doycheva, & Wentura, 2012), mostly by comparing the processing of L1 and L2 by the same individual and occasionally by comparing native and non-native speakers' processing of the same language (e.g., Eilola & Havelka, 2011).

The current study extends this line of research to Cantonese–Mandarin bilinguals. As discussed above, previous studies are highly focused on European languages, with at least one (usually English) and often two European languages in the language pair. When the two languages are typologically close (e.g., Catalan-Spanish), cognates – i.e., translation equivalents with similar orthographic and phonological forms – are eschewed, presumably to avoid mixing cognates and non-cognates in the stimuli and to prevent possible confusion about which language is being presented. To this end, the study of Cantonese–Mandarin bilinguals not only extends the research to non-European languages, but also provides a precious opportunity of examining ORTHOGRAPHICALLY IDENTICAL BUT PHONOLOGICALLY DIFFERENT AFFECTIVE COGNATES, which exist in large quantities between Cantonese and Mandarin, due to the high degree of linguistic and cultural continuity among Chinese languages. The overlap between the Cantonese and Mandarin lexicons allows us to tease apart critical factors that are largely confluent in previous research, such as the roles of lemma and lexeme and the influence of lexical and cultural factors. In the following, we will first review in more detail the mediating factors and underlying mechanisms of differential affective processing, before introducing the design of the current study.

Despite the wide presence of supporting evidence, multiple studies reported a lack of differential affective processing (Altarriba & Basnight-Brown, 2012; Ferré et al., 2013; Min & Schirmer, 2011) – or even a reversed L1-L2 difference (i.e., L2 being more emotional than L1, see Marian & Kaushanskaya, 2008; Ong, Hussain, Chow, & Thompson, 2017). These exceptions can be linked to the discussion of what factors modulate the presence (or absence) and magnitude of differential affective processing. There are at least three sources of modulating factors: (1) the bilingual's L2 proficiency and experience, (2) the properties of affective words, and (3) the nature of the task and measure. Overall, bilinguals with higher L2 proficiency (i.e., more balanced bilinguals), who typically have earlier age of acquisition and higher levels of immersion and usage frequency of the L2, are less likely to show differential affective processing (Ayçiçegi-Dinn & Caldwell-Harris, 2009; Degner et al., 2012; Eilola et al., 2007; Ferré et al., 2010, 2013; Harris, Gleason, & Ayçiçeği, 2006; Sutton et al., 2007). Within the group of affective words, taboo words and other negative words tend to elicit stronger emotionality effects than positive words, and so do emotion words compared to emotion-laden words. Correspondingly, these variations translate into varying degrees of L1-L2 differences, with larger L1-L2 differences observed for strongly negative words (compared to positive words) and emotion words (compared to emotion-laden words) (Altarriba & Basnight-Brown, 2012; Ayçiçegi-Dinn & Caldwell-Harris, 2009; Eilola & Havelka, 2011; Garrido & Prada, 2018; Kazanas & Altarriba, 2016). To this end, some researchers claim that the valence of negative words is greatly subdued or suppressed in the L2, causing less emotional disruption to the bilingual speaker than the L1 counterparts would (Jończyk, Boutonnet, Musiał, Hoemann, & Thierry, 2016; Marian & Kaushanskaya, 2008; Wu & Thierry, 2012; see more discussion below). Finally, the experimental task (or measure) used to gauge the emotionality effect may also play a role. In general, physiological measures (skin conductance responses, pupil dilation) tend to be more sensitive than behavioral measures (e.g., reaction times) or subjective emotionality ratings. In fact, a number of studies reported L1-L2 differences ONLY in physiological measures while no differences were found in behavioral measures or ratings (Caldwell-Harris et al., 2011; Eilola & Havelka, 2011; Iacozza et al., 2017; Ponari et al., 2015; Segalowitz et al., 2008).

The phenomenon of differential affective processing can be imputed to cross-language differences in terms of learning context, the formation of emotional memory, and the structure of the affective lexicon. For sequential bilinguals, the L1 is learned through immersive experiences since birth whereas the L2 is usually learned through classroom instruction with impoverished emotional context. As a result, bilinguals are much more likely to form emotional memories associated with the L1 than with the L2 (Harris et al., 2006; Robinson & Altarriba, 2018). The importance of emotional embodiment is reflected in the special status of childhood reprimands and endearments, which often show the highest amount of L1-L2 difference among affective words, even for early bilinguals (i.e., heritage speakers, see Harris et al., 2006). It has been argued that without being grounded in emotional context, L2 affective words are less likely to develop a connection with the corresponding emotional

[1]Throughout the paper, bilingual backgrounds are denoted in a hyphenated format of *L1-L2*, with L1 preceding L2.

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

concepts; instead, they may only be connected to the L1 counterparts at the lexical level (Altarriba, 2001, 2003). This view is compatible with the revised hierarchical model of the bilingual lexicon (Kroll & de Groot, 1997; Kroll & Stewart, 1994), which separates conceptual representations (i.e., meaning) from lexical representations. The model assumes that novice L2 learners tend to acquire L2 words via lexical mediation by translation equivalents in the L1, and that direct conceptual links between meanings and L2 words will only be established as the learner obtains higher proficiency and more experience with the L2. Thus, the asymmetry between L1 and L2 conceptual links will give rise to differential affective processing, as L2 affective words have less direct/automatic access to the emotional concepts.

The L1-L2 difference goes beyond single word/phrase processing, as bilinguals also tend to be more rational and less emotional in higher-level decision-making, e.g., when faced with moral dilemmas or financial incentive calculation (i.e., the "foreign-language effect"; Costa et al., 2017, 2014; Gao, Zika, Rogers, & Thierry, 2015; Keysar, Hayakawa, & An, 2012; Pavlenko, 2017; Stankovic, Biedermann, & Hamamura, 2022; but see also Oganian, Korn, & Heekeren, 2016). The emotional distance manifested in L2 processing also explains why bilinguals are more at ease with describing harsh or embarrassing memories in the L2 (Altarriba & Santiago-Rivera, 1994; Bond & Lai, 1986; Dewaele & Pavlenko, 2002; Heredia & Altarriba, 2001; Marian & Kaushanskaya, 2008; Wu & Thierry, 2012) or using affectionate language in the L2 if such practice is not encouraged in the native culture (Caldwell-Harris et al., 2011).

A common criticism of the existing literature of bilingual affective word processing is the lack of attention to crosslinguistic cultural and lexical differences (Grosjean, 1998; Pavlenko, 1999). The wide use of translation equivalents in previous studies assumes a crosslinguistic equivalence of both emotional concepts and affective expressions, although arguments for the opposite have been made repeatedly (Altarriba, 2003; Pavlenko, 2008; Robinson & Altarriba, 2018). Anyone who has attempted to translate affective words and phrases from one language to another would agree that the notion of a precise, one-to-one mapping of translation equivalents is nonexistent in most cases. For example, the Chinese expression 幸災樂禍 'pleasure derived from someone else's misfortune' is not readily available in English, although German has a similar expression *schadenfreude*. Let alone nuanced differences in intensity and valence between translation equivalents. Given such, how can we be sure that any observation of differential affective processing is due to how bilinguals process L1 and L2 words *per se*, but not underlying differences between the L1 and L2 words?

The current study intends to fill this gap by focusing on two Chinese languages, Hong Kong Cantonese ("Cantonese" or "CAN" hereafter) and Standard Mandarin ("Mandarin" or "MAN" hereafter), which share a highly overlapping lexicon. While Mandarin is often the *lingua franca* of a Chinese-speaking community, Cantonese is the most common native language of the population of Hong Kong and one of the major languages in the Canton province of Mainland China. As with other varieties of Chinese, most (but not all) of the words in Cantonese and Mandarin belong to the lexicon of Standard Written Chinese, with no distinction in orthography due to the logographic nature of the Chinese writing system but mutually unintelligible in terms of the spoken forms (although phonologically often partially overlapping).[2] Figure 1 shows the lexical model of a typical affective cognate shared between Cantonese and Mandarin: the lemma of 憤怒 'rage', which contains the syntactic and semantic properties of the word, is shared by Cantonese and Mandarin, and the conceptual representation of rage is shared as well; however, at the lexeme level, the word has different phonological forms in the two languages.[3] For Cantonese–Mandarin bilinguals, most of the effort when acquiring the nonnative Mandarin lexicon is spent on learning the pronunciations of words that have already been acquired in native Cantonese.

The critical question we ask in this study is whether we still find evidence for differential affective processing, given that the Cantonese–Mandarin cognates share the lemmas and only differ in phonological forms. In other words, will Cantonese–Mandarin bilinguals show greater emotional arousal when they are exposed to the Cantonese pronunciation of an affective word compared to the Mandarin pronunciation of the same word? The measure we use to gauge emotional arousal is pupil dilation (i.e., enlargement of pupil size), which has long been associated with a multitude of physical, psychological, and environmental factors such as illuminance, fatigue, sexual and emotional arousal, listening and processing effort, etc. (Hartmann & Fischer, 2014; Knapen et al., 2016; Koelewijn, Zekveld, Festen, & Kramer, 2012; Ksiazek, Wendt, Alickovic, & Lunne, 2018; Partala & Surakka, 2003; Robison & Unsworth, 2019; Tamási, McKean, Gafos, Fritzsche, & Höhle, 2017; Tryon, 1975; Wagner, Nagels, Toffanin, Opie, & Başkent, 2019; Winn, Wendt, Koelewijn, & Kuchinsky, 2018). Despite the general consensus that the pupil dilates in response to emotionally charged stimuli (Bradley, Miccoli, Escrig, & Lang, 2008; Hess, 1965; Janisse, 1974), mixed findings have been reported for emotionality effects on pupil dilation during word processing. Bayer et al. (2011) actually found smaller – instead of larger – dilations for high-arousing words than low-arousing words; Võ et al. (2008) reported smaller dilations for negative than for positive words, which is in line with Bayer and colleagues' finding if we assume that negative words tend to be more arousing than positive words. On the other hand, Kuchinke, Võ, Hofmann, and Jacobs (2007) reported null effects of word valence on pupil dilation. A potential reason that underlies the mixed findings is the mediation of processing effort. Previous studies have shown that pupil dilations may also reflect enhanced lexical processing effort, e.g., low-frequency words, which require more retrieval effort, tend to elicit higher pupil dilations than high-frequency words (Kuchinke et al., 2007; Schmidtke, 2014). As mentioned above, words with high emotionality are associated with facilitated cognitive processing, evidenced by faster responses in lexical decision and identification, and this facilitative effect may lead to the reduction or cancellation of the anticipated emotionality effects on pupil dilation. Despite the potential interference of processing effort, a more recent study by Iacozza et al. (2017) demonstrated that pupil dilation is sensitive enough to differentiate emotional arousal during L1 vs. L2 processing: greater pupillary responses were recorded

[2]Strictly speaking, there are two Chinese writing systems, traditional Chinese and simplified Chinese, both of which are logographic. The distribution of the two systems is separated by geography (e.g., simplified Chinese in Mainland China, and traditional Chinese in Taiwan and Hong Kong), not by the spoken variety (e.g., Mandarin vs. Cantonese).

[3]Cai, Pickering, Yan, & Branigan (2011) proposed that Cantonese and Mandarin dative verbs have separate lemmas in the bilingual lexicon, based on evidence from structural priming, but as far as we know, there is no evidence to argue for separate lemmas between Cantonese and Mandarin in other word categories.

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

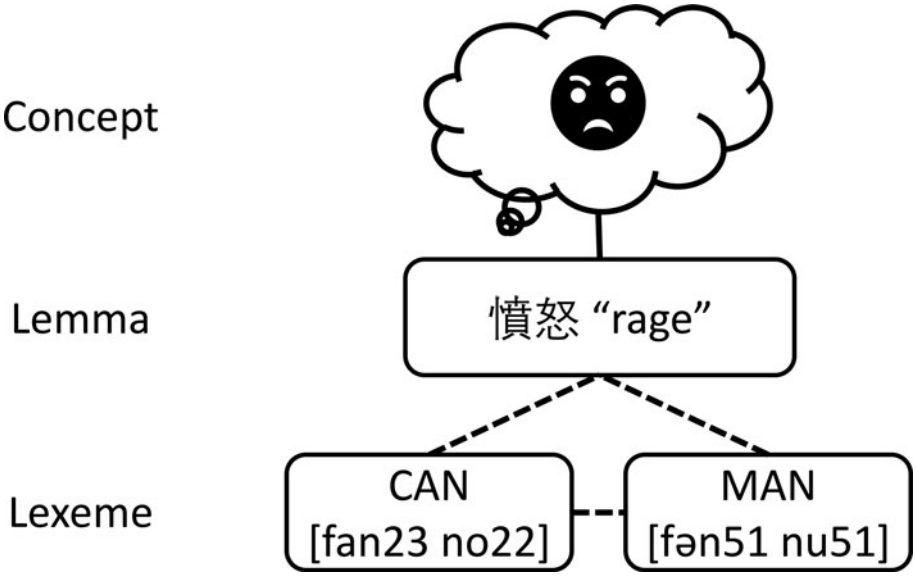


Fig. 1. Representations of a Chinese affective word, 憤怒 'rage' at the conceptual, lemma, and lexeme levels. Phonological forms in Cantonese (CAN) and Mandarin (MAN) are transcribed in the IPA, with 5-level tone transcription (1 = low, 5 = high). According to Yao and Chang (2016), lexical-phonetic connections may also exist between the phonological forms of a cognate in two Chinese languages, with the strength of the connection dependent on the degree of crosslinguistic phonological similarity, the bilingual's language experience, etc.

when participants read out aloud emotional texts in their L1 than when the same task was performed in the L2, although L2 processing should require greater effort.

In the current study, we record the pupil sizes of Cantonese–Mandarin bilinguals when listening to affective words in Cantonese (L1) or Mandarin (L2) pronunciations. The main hypothesis is that greater pupil dilations will be observed when Cantonese pronunciations are presented. Speakers with less balanced bilingual skills (i.e., weaker L2 proficiency) are predicted to show greater differential affective processing than more balanced bilinguals. We also predict a stronger effect of differential affective processing for strongly negative words.

Methods

Participants

52 participants (18–25 y.o.; 42F, 10M) were recruited from a university in Hong Kong. All the participants were native speakers of Cantonese, with Cantonese-speaking parents, born in Hong Kong (N = 48) or moved to Hong Kong between the age of one and four (N = 4). The participants started to learn Mandarin at a young age (M = 5.31 y.o., SD = 1.63), mostly in preschools, and had on average 1.82 hours per week (SD = 1.52) of Mandarin classes or classes taught in Mandarin in primary and secondary schools. Most of the participants rated their Mandarin proficiency as Intermediate (N = 21) or Upper Intermediate (N = 13), whom we refer to as Mid-Proficiency bilinguals, in contrast with the rest of the participants who self-rated as Beginner/Elementary (i.e., Low-Proficiency; N = 8) or Advanced/Proficient (i.e., High-Proficiency; N = 10).[4] Given the "biliterate and trilingual" policy in the education system of Hong Kong, all the participants also had early and continuing exposure to English, and a few participants reported knowledge of additional languages such as Japanese, Korean and French. Two participants are left-handed, one ambidextrous, and the rest right-handed.

A separate group of 16 non-Chinese-speaking listeners were recruited to participate in the validation test of the auditory stimuli.

Stimuli

The word list consists of 130 disyllabic words, of which roughly half are affective words (N = 70) and the rest neutral fillers (N = 60) (see Table S1 for the complete word list). The affective words are further divided into taboo words (N = 10), emotion words (N = 30) and emotion-laden words (N = 30). According to Janschewitz (2008), taboo words can be divided into several subcategories, including religious profanity, cultural taboos (slangs or clinical terms referring to sexuality, bodily functions, death, etc.), and verbal insults based on a person's race, gender, sexuality. The taboo words in this study belong to either cultural taboos (e.g., 性交 'intercourse', 大便 'feces', 死亡 'death') or verbal insults (e.g., 賤人'bitch', 畜生 'bastard').[5] The group of emotion words we used are chosen from the Chinese emotion word database compiled by Lin and Yao (2016), which also reports ratings of emotion category, intensity and valence for the written forms of the words from Chinese speakers in mainland China, Hong Kong, and Singapore. We chose six words in each basic emotion category (HAPPINESS, SADNESS, ANGER, FEAR, and SURPRISE): three words from the top intensity range and the other three from the bottom intensity range, with preference given to words that elicited similar intensity ratings from mainland Chinese and Hong Kong speakers. Valence ratings are largely aligned with emotional categories (e.g., HAPPINESS – positive, SADNESS/ANGER/FEAR – negative, SURPRISE – neutral). The emotion-laden words (15 negative, 15 positive) and neutral fillers (e.g., 圓圈 'circle') are both randomly sampled from the word lists in Chen et al. (2015). In subsequent data analysis, two test words (one emotion word and one emotion-laden word) are excluded because they were accidentally included in the practice trials (see more in

[4]In addition to the 52 participants reported here, another nine participants also completed the experiment, but their data were excluded from the analysis because they arrived in Hong Kong after the age of 12 (N = 3), had a history of speech/language disorder (N = 1), had extensive Mandarin input (>10hrs per week in primary/secondary schools; N = 4), or produced an alarmingly high percentage of erroneous trials with no valid measurements (>50%; N = 1).

[5]One of the taboo words in this study, 去死 'go to die', is arguably both a cultural taboo and a verbal insult

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

Data preprocessing), leaving 128 test words (68 affective words, 60 neutral fillers) in the analysis.

Audio recordings of the words were made in a sound-treated booth with two female speakers, a native Cantonese speaker from Hong Kong (reading in Cantonese) and a native Mandarin speaker from northern China (reading in Mandarin), both in their 20s and naïve to the purpose of the study. Individual word recordings were scaled to an average intensity of 65 dB. Overall, the Cantonese word productions were slightly longer than the Mandarin productions (M_{CAN} = 833 ms, SD_{CAN} = 116, M_{MAN} = 729 ms, SD_{MAN} = 73; $p < 0.001$ in paired *t*-test). To ascertain that the speakers maintained a neutral tone when producing the stimuli, a validation test was conducted, where a separate group of 16 listeners – all of whom were non-speakers of either Cantonese or Mandarin and had no knowledge about the main study – listened to the recordings and provided intuitive judgements about the emotional intensity they perceived from the recordings on a 5-point scale. A cumulative link mixed model on intensity rating, fitted with the `clmm()` function of the *ordinal* package (Christensen, 2019; version 2019.12-10) in R (R Team, 2019; version 3.6.1), showed no significant effects of Language (CAN, MAN), WordType (taboo, emotion words, emotion-laden words, neutral filler), or the interaction of the two (all *p*s > 0.1), while controlling for by-listener and by-item random intercepts, by-listener random slopes for Language and WordType, and by-item random slope for Language.

Although the test words are shared between Cantonese and Mandarin lexicons, their use frequency may differ across languages. To control for frequency effects, we used subjective frequency ratings provided by the current participants (see ***Procedure*** below) instead of corpus-based frequency counts, due to the lack of corpus resources for Cantonese and the potential problem of estimating usage frequency in a nonnative language (Mandarin) with counts from monolingual corpora. Overall, the Cantonese pronunciations of the test words were rated as more frequently heard (Mean = 4.49 on a 7-point scale) than the Mandarin counterparts (Mean = 4.14; $p < .001$ in paired t-test), although the two ratings are closely correlated (r = 0.70). Participants also gave more "unknown" responses to the Mandarin stimuli (N_type = 81; N_token = 201) than the Cantonese counterparts (N_type = 27; N_token = 64).

Procedure

The experiment was run in SR Research Experiment Builder (SR Research, 2011). The participant was seated in front of a desktop computer in a soundproof booth, with the head position stabilized by a desk-mounted chin and forehead rest at a horizontal distance of 78 cm from the computer screen. The sound stimuli were played through a set of AKG K77 headphones, and the participant's pupil responses were tracked by an EyeLink 1000 eye tracker. The participant was told that they would be hearing and evaluating the usage frequency of a series of Cantonese or Mandarin words. Each session consisted of eight practice trials and 130 test trials (one for each test word, in a randomized order). Each participant would hear half of the test words in Cantonese, and the other half in Mandarin, presented in two language-specific blocks. The participant could take a short break between the two blocks, and a complete eye-tracking session lasted about 40 minutes. The assignment of languages and the order of language blocks were balanced across participants. Each trial would begin with a fixation period of 1.5 s, with a cross sign (+) that is 60 pixels in height and width shown at the center of the screen. After the stimulus started playing, the cross sign remained on the screen for a period of 2–5 s, to allow enough time for observing stimulus-evoked pupil size changes. The duration of this period (Mean = 2762 ms, SD = 646) varied based on the length of the stimulus and a quasi-random fluctuation. The participant was instructed to always look at the cross sign when it was on the screen. After the cross sign disappeared, the participant would press a key to proceed to rate how often they heard the word on a 7-point scale (7 = frequently, 1 = rarely, 0 = unknown word). An interval of 2 s was administered between trials in the same block. Throughout the experiment, there was no mention in any form of the emotionality of the stimuli. After the eye-tracking experiment, the participant would complete a survey about their language background and their experience with both Cantonese and Mandarin.

Data preprocessing

In each experimental trial, the participant's right pupil size and gaze coordinates were tracked at a sampling rate of 1000 Hz by the eye tracker. The tracking period – during which a cross sign remained on the screen – can be divided into two consecutive intervals: the fixation period (before stimulus onset) and the stimulus tracking period (after stimulus onset and before the beginning of the frequency rating task).

Following the recommendation in previous studies (Mathôt, Fabius, Van Heusden, & Van Der Stigchel, 2018; Schacht, Dimigen, & Sommer, 2010), raw pupillometry data were preprocessed with a pipelined procedure that (1) replaced intervals of blinks, saccades, and outlier measurements (2.5 standard deviations away from the mean pupil size of the current trial) with interpolations of nearest valid values, (2) smoothed the measurements with a floating 5-point window, (3) down-sampled the signal from 1000 Hz to 10 Hz (i.e., 10 samples per second) by averaging, and (4) baseline-corrected the measurements in the stimulus tracking period by subtracting the mean pupil diameter during the last 200 ms of the pre-stimulus fixation period. The result of data preprocessing is a time series of task-evoked pupil diameter measurements for each trial. Furthermore, we removed test trials that have (1) more than 50% of measurements replaced by interpolations (i.e., suggesting anomaly in pupil size measurement) or (2) more than 50% of the measurements that occurred outside of an area of 300 pixels in diameter around the cross sign (i.e., suggesting extensive eye movement; Mathôt et al., 2018). Finally, we also excluded the test trials of two critical words that were accidentally included in the practice trials, as well as trials that yielded "unknown" responses in the frequency rating task (i.e., the participant did not recognize the word). Altogether 25.7% of the test trials were excluded, leaving 5,021 trials (featuring 52 participants and 128 critical words spoken in two languages) and 132,173 pupil size measurements for analysis.

Statistical analysis

In the rest of the paper, unless otherwise specified, all the pupillometry analysis is based on task-evoked pupil diameter measurements. The analysis focuses on a 3000 ms window after stimulus onset, which covers the whole durations of most of the trials and contains the critical time windows reported in previous

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

pupillometry studies with similar stimuli or trial durations (e.g., 0–1200 ms in Schmidtke, 2014; 0–1500ms in Kuchinke et al., 2007; 1000 ms–1500 ms in Bayer et al., 2011 and Schacht et al., 2010; 1000–2000 ms in Tamási et al., 2017). Generalized additive mixed modeling (GAMM; Sóskuthy, 2017, 2021; van Rij, Hendriks, van Rijn, Baayen, & Wood, 2019; Wieling, 2018; Wood, 2017) is used to model whether the shape of the pupil diameter curve (over time) varies by language, word type, Mandarin proficiency, and their interactions. All the models are built with the `bam()` function in the *mgcv* package (Wood, 2011, 2017; v1.8.38) in R, while model checking and data visualizations are conducted with functions from the *itsadug* package (van Rij, Wieling, Baayen, & van Rijn, 2017; v2.4).

The GAMM technique effectively unveils both linear and non-linear dependencies – fixed and random – in the data while keeping under control the inherent autocorrelation in time series observations, an issue that has been shown to affect other statistical methods of analyzing time series data such as the growth curve analysis (Huang & Snedeker, 2020). All the GAMMs reported in this paper incorporate an autoregressive error model at lag = 1 (AR(1)), with an autocorrelation coefficient *rho* estimated from a maximally similar base model without AR(1). Autocorrelations of model residuals in the final models are around 0.3 at lag = 1.[6] Model fits are checked with `gam.check()`. Initial model testing showed residual distributions with heavy tails on both ends, so a scaled-*t* link was used in the final models (Wieling, 2018). The values of basis dimension (*k*) are also adjusted according to the results from model fit testing. When a *k* value is too low, it would be doubled until the issue disappears or when *k* reaches the maximum (i.e., the number of unique values of the variable). All the models reported below have sufficient *k* values.

Results

Data summary

Overall, baseline pupil diameter shows a general trend of decreasing over time within a testing block (see Figure S2), possibly due to fatigue or disengagement, as noted in previous studies (McGinley, David, & McCormick, 2015; Winn et al., 2018). Against this backdrop, task-evoked pupil diameter in the analysis window is in general greater than zero, suggesting overall pupil dilation (compared to the baseline) after stimulus onset, although the degree of dilation tends to be smaller in the 2nd block than in the 1st block, probably also due to fatigue (see Figure S3). The peak of dilation usually occurs at around $t = 1500$ ms, at which point the increase in pupil diameter is on average 8.34% (SD = 6.57%) from the baseline diameter.

Modeling task-evoked pupil diameter change over time with GAMMs

The dependent measure in all the GAMMs is task-evoked pupil diameter. To control for the effects of fatigue, gaze position (Gagl, Hawelka, & Hutzler, 2011), and usage frequency on pupil size, all the models contain smooths for a group of control predictors including block (IsBlock2; a binary variable that is 1 if the response is from Block 2, 0 if from Block 1), trial order within a block (TrialNO), gaze coordinates (Gaze_x, Gaze_y), and subjective frequency rating (SubjectiveFreq). The models also include random smooths for participant (SubjID) and word (WordID). More complex random structures (e.g., random intercepts or slopes for certain predictors, or random smooths for individual trials) are not included due to model non-convergence.

In line with the central inquiries, the analysis is conducted in two stages: Level-1 models have language × word type as the critical predictors for the investigation of differential affective processing effect, and Level-2 models further examine how Mandarin proficiency modulates differential affective processing (i.e., language × word type × proficiency).

In the first stage, we first built a model `gamm.1a.AR`, where the interaction of language and word type was coded as a categorical interactional variable LangWtype, which combines the levels of language ("M" for Mandarin, "C" for Cantonese) and word type ("NT" for neutral filler, "EL" for emotion-laden word, "EW" for emotion word, "TB" for taboo), with "M.NT" (i.e., Mandarin neutral filler) being the reference level. Model formula and result summary can be found in Appendix S4.

All the control predictors in `gamm.1a.AR` have significant smooth terms (see Figure S5 for the visualization of the control effects). Specifically, later trials (either in Block 2 or occurring later within a block) and trials with more frequently heard stimuli tend to have reduced pupil dilations, which is compatible with the predicted effects of fatigue and frequency.

Regarding the critical predictor of LangWtype, the parametric effect shows that both Mandarin taboos (M.TB: $\beta = 53.16$, $p < .001$) and Cantonese taboos (C.TB: $\beta = 74.61$, $p < .001$) have significantly higher curves than the baseline Mandarin neutral fillers, but no difference in intercept between other levels of LangWtype and the baseline (all ps > .05). Among the smooth terms regarding LangWtype, only the smooth of Cantonese taboos is significantly different from zero ($p < .001$). As shown in Figure 2, the overall predicted response curve is higher for taboos than neutral fillers in both languages, from slightly after 1 s to the end of the window.

To verify the significance of the taboo vs. neutral difference and to compare the size of the taboo vs. neutral difference across languages, we built `gamm.1b.AR`, where the critical interaction of language and word type was re-coded as seven binary variables (IsCantonese, IsEL, IsEW, IsTB, IsC.EL, IsC.EW, IsC.TB) defined as follows (see Appendix S6 for model formula and result summary):

- IsCantonese = 1 if language is Cantonese, 0 otherwise.
- IsEL = 1 if word type is emotion-laden word, 0 otherwise.
- IsEW = 1 if word type is emotion word, 0 otherwise.
- IsTB = 1 if word type is taboo, 0 otherwise.
- IsC.EL = 1 if word type is emotion-laden word and language is Cantonese, 0 otherwise.
- IsC.EW = 1 if word type is emotion word and language is Cantonese, 0 otherwise.
- IsC.TB = 1 if word type is taboo and language is Cantonese, 0 otherwise.

Since the smooths of the binary variables include the differences in both the intercept and non-linear terms, there are no

[6]An autocorrelation coefficient of 0.2 or below usually indicates that the autocorrelation issue is sufficiently alleviated. In view of the slightly higher autocorrelation coefficient reported here, we replicated each analysis presented below with a halved dataset (i.e., with only the odd-numbered observations in each time series) to assess the reliability of the modeling results. The halved dataset has lower inherent autocorrelation since the observations are farther apart. The resulting GAMMs based on the halved datasets have an autocorrelation coefficient of around 0.18, with all the fixed effects unshifted from the full models.

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

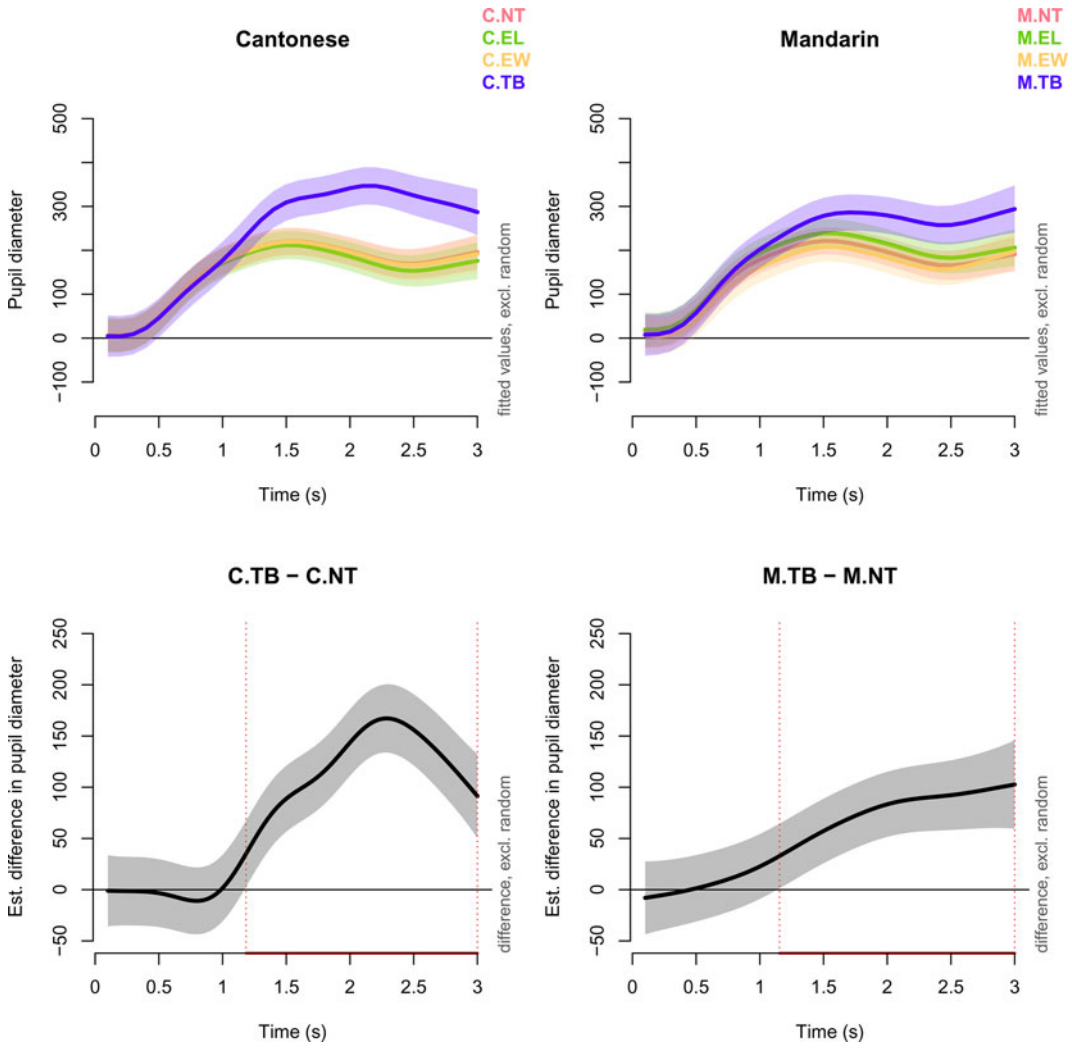


Fig. 2. Fitted pupillary response smooths for each level of LangWtype and visualization of the difference between taboos and neutral fillers by language (left column: Cantonese; right column: Mandarin), based on `gamm.1a.AR`. The shaded bands denote the pointwise 95%-confidence intervals.

parametric terms for the binary variables. Given the coding scheme above, the reference smooth represents Mandarin neutral fillers, and all other Language × WordType levels are modeled by additional binary smooths. Specifically, Cantonese neutral fillers are represented by `s(Time, by = IsCantonese)`; Mandarin taboos are presented by `s(Time, by = IsTB)`; Cantonese taboos are represented by `s(Time, by = IsCantonese) + s(Time, by = IsTB) + s(Time, by = IsC.TB)`.

Model summary shows that among the seven newly added binary variables, only IsTB and IsC.TB have smooths that are significantly different from zero ($p < .05$). The visualization (see Figure S7) shows that both IsTB and IsC.TB are associated with higher response curves, from around 1.2 s to the end of the window for IsTB and between 2 s and 2.5 s for IsC.TB. These effects confirm that task-evoked pupil diameter is significantly larger for taboo words than for neutral words in both languages, and that the effect is even greater in Cantonese than in Mandarin.

To summarize, the results of Level-1 models (`gamm.1a.AR` and `gamm.1b.AR`) provide evidence for differential affective processing in Cantonese–Mandarin bilinguals, with the largest emotionality effects observed for taboo words. While bilingual listeners exhibit higher pupil dilation to taboo words presented in both languages (compared with neutral fillers), the magnitude of the emotionality effect is stronger when the words are presented in L1 Cantonese pronunciations than in L2 Mandarin pronunciations.

Next, we added Mandarin proficiency to the model and its interaction with language and word type, in order to investigate whether differential affective processing is modulated by the bilingual's proficiency of the L2. Since the non-taboo affective words did not show significant differences from neutral fillers in the Level-1 models, we focused on only taboos and neutral fillers in the Level-2 models. The dataset includes 72,042 pupil size measurements from 2,739 trials. We first built `gamm.2a.AR`, with a categorical interactional variable LangWtypeMprof, which combines all the levels of language, word type, and Mandarin proficiency ("LowP" for low proficiency, "MidP" for intermediate proficiency, and "HighP" for high proficiency), with M.NT.LowP (i.e., low-proficiency bilinguals' responses to Mandarin neutral fillers) as the reference level. Model summary (see Appendix S8) shows that among the levels of LangWtypeMprof, only C.TB.LowP (i.e., low-proficiency bilinguals' responses to Cantonese taboos) has a significant parametric term – in the positive direction, but all the levels are associated with significant

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

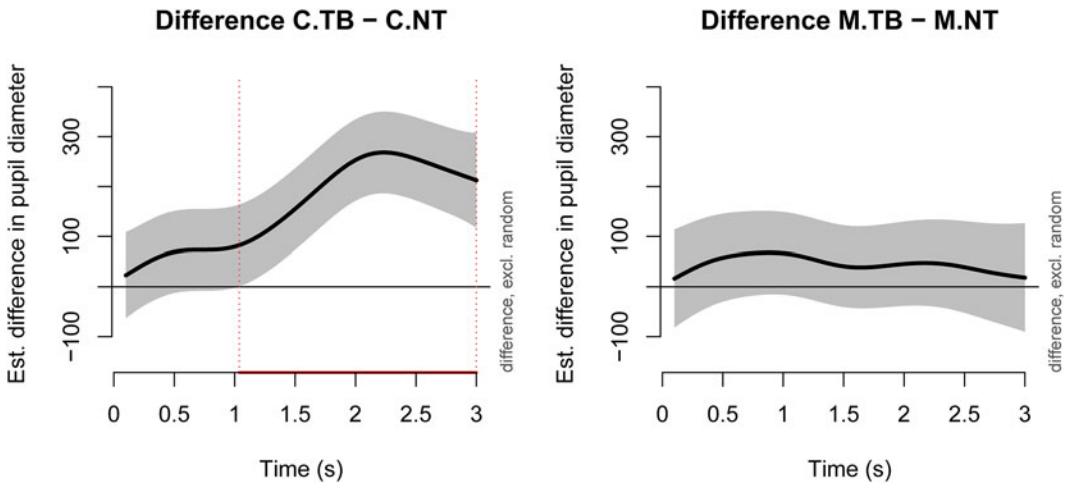


Fig. 3. Visualization of the difference between the fitted pupillary response smooths for taboos and neutral fillers in Cantonese (left) and Mandarin (right) for LOW-PROFICIENCY BILINGUALS in `gamm.2a.AR`. The shaded bands denote the pointwise 95%-confidence intervals.

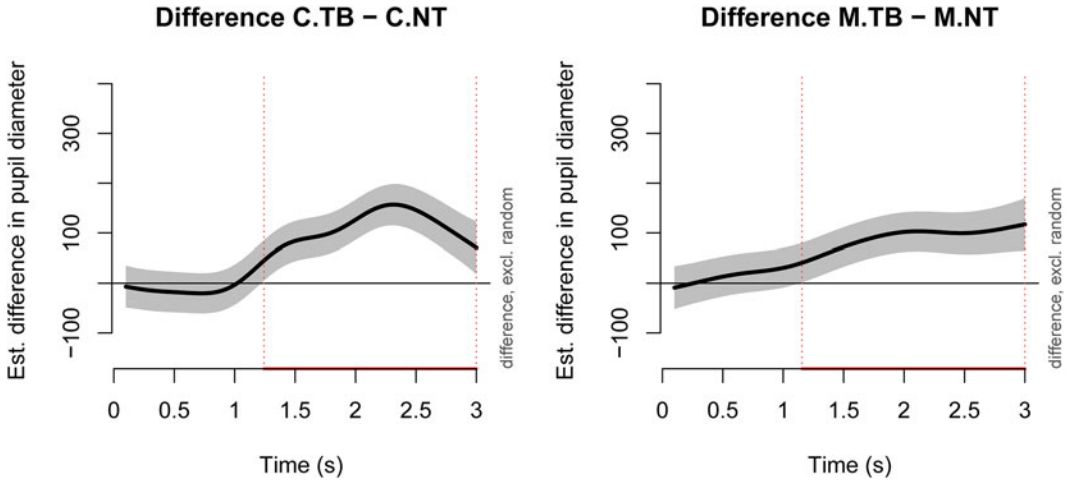


Fig. 4. Visualization of the difference between the fitted pupillary response smooths for taboos and neutral fillers in Cantonese (left) and Mandarin (right) for MID-PROFICIENCY BILINGUALS in `gamm.2a.AR`. The shaded bands denote the pointwise 95%-confidence intervals.

smooths over time. The visualization in Figure 3 reveals that low-proficiency bilinguals' response smooth for Cantonese taboos is higher than that of Cantonese neutral fillers, but no such difference exists in Mandarin. By contrast, both mid- and high-proficiency bilinguals have higher response curves for taboos than neutral fillers in both languages (Figure 4 and Figure 5). The taboo-neutral difference, if existent, is usually evident in the second half of the window, starting between 1 s and 1.5 s.

To confirm the significance of the effects observed in `gamm.2a.AR`, we fitted `gamm.2b.AR`, which models the language × word type × proficiency interaction with a categorical variable of Mandarin proficiency (Mprof) and three binary variables for the language × word type interaction for each proficiency group (see Appendix S9). As an example, the binary variables for the low-proficiency group are as follows:

- IsTB.LowP = 1 if word type is taboo and Mandarin proficiency is low, 0 otherwise.
- IsC.LowP = 1 if language is Cantonese and Mandarin proficiency is low, 0 otherwise.
- IsC.TB.LowP = 1 if language is Cantonese, word type is taboo and Mandarin proficiency is low, 0 otherwise.

Accordingly, low-proficiency bilinguals' response smooths are modeled by the following parametric and smooth terms (excluding control factors and random effects):

- Mandarin neutral fillers: MprofLowP + s(Time):MprofLowP
- Mandarin taboos: MprofLowP + s(Time):MprofLowP + s(Time, by= IsTB.LowP)
- Cantonese neutral fillers: MprofLowP + s(Time):MprofLowP + s(Time, by= IsC.LowP)
- Cantonese taboos: MprofLowP + s(Time):MprofLowP + s(Time, by= IsTB.LowP) + s(Time, by= IsC.LowP) + s(Time, by= IsC.TB.LowP)

If low-proficiency bilinguals show a significant effect of differential affective processing, IsC.TB.LowP should show a significant, positive smooth over time. Similar predictions can be made for mid- and high-proficiency bilinguals.

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

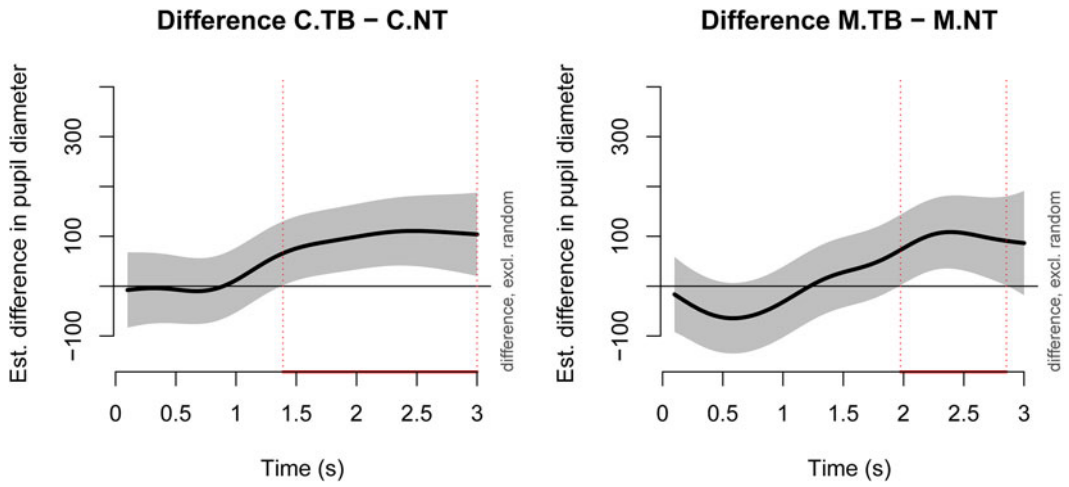


Fig. 5. Visualization of the difference between the fitted pupillary response smooths for taboos and neutral fillers in Cantonese (left) and Mandarin (right) for HIGH-PROFICIENCY BILINGUALS in `gamm.2a.AR`. The shaded bands denote the pointwise 95%-confidence intervals.

The results of `gamm.2b.AR` show no cross-language difference in neutral fillers for any proficiency group (`s(Time):IsC.LowP`, `s(Time):IsC.MidP`, and `s(Time):IsC.HighP` are non-significant; all *p*s > .05). But regarding the taboo-neutral difference, the three proficiency groups show different patterns. Low-proficiency bilinguals have significant difference between neutral fillers and taboos in Cantonese (`s(Time):IsC.TB.LowP`; $p = .002$) but not in Mandarin (`s(Time):IsTB.LowP`; $p = .71$). Mid-proficiency bilinguals show a significant taboo-neutral difference in Mandarin (`s(Time):IsTB.MidP`; $p < .001$) and an additional cross-language difference for the taboo-neutral comparison in Cantonese (`s(Time):IsC.TB.MidP`; $p < .001$). High-proficiency bilinguals also show a significant taboo-neutral difference in Mandarin (`s(Time):IsTB.HighP`; $p < .001$), but no additional cross-language difference for the taboo-neutral comparison in Cantonese (`s(Time):IsC.TB.HighP`; $p = .8$), suggesting that the same amount of taboo-neutral difference is present in both languages.

The visualization of the partial effects of the binary variables in `gamm.2b.AR` (Figure S10) confirms that the directions of the significant effects are all in the predicted, positive direction. It can also be seen that the effect size of IsC.TB.LowP is much larger than that of IsC.TB.MidP, confirming that the magnitude of the language × word type interaction is greater for low-proficiency bilinguals.

Taken together, the results of Level-2 models reveal a continuum of the degree of differential affective processing, with low-proficiency bilinguals on the highest end of the continuum, showing taboo-neutral differences only in Cantonese (L1) but not in Mandarin (L2), and high-proficiency bilinguals on the lowest end, showing taboo-neutral differences in both languages to the same degree. Mid-proficiency bilinguals are in the middle, showing taboo-neutral differences in both languages but with a slightly stronger effect in Cantonese than in Mandarin.

Discussion

An important feature of the current study is the focus on cognates, which were eschewed in previous research on differential affective processing. All the test words used in the current study are identical between Cantonese and Mandarin in both orthography and meaning, differing only (partially) in the phonological form. The existing literature has consistently found evidence for the so-called cognate effect (also known as the cognate advantage) in bilingual/L2 processing: compared to non-cognates, cognates are easier to recognize and name (Costa, Santesteban, & Caño, 2005; Dijkstra, Hilberink-Schulpen, & Van Heuven, 2010; Lalor & Kirsner, 2000; Peeters, Dijkstra, & Grainger, 2013), more susceptible to phonetic transfer and more likely to be influenced by the crosslinguistic counterparts (Amengual, 2012; Brown and Harper, 2009; Flege & Munro, 1994; Goldrick, Runnqvist, & Costa, 2014; Simonet & Amengual, 2020; Yao & Chang, 2016; but see Cochrane, 1980; Flege, Frieda, Walley, & Randazza, 1998; Flege, Takagi, & Mann, 1995 for different views). Given such, cognates should be LESS likely to elicit differential affective processing compared to non-cognates. Thus, the fact that we still observe significant cross-language differences with Cantonese–Mandarin affective cognates attests to the resilience of the differential affective processing phenomenon under high linguistic similarity. This study also contributes new data from Chinese languages to the research of cognate processing, which is disproportionately concentrated on European languages. The cognates shared between Cantonese and Mandarin (as well as other Chinese varieties) are uniquely characterized by their vast amount, identical orthography, and often mutually unintelligible pronunciations due to the differences in the segmental/tonal inventory across languages.

Locus of the emotionality effect

For differential affective processing to occur between Cantonese and Mandarin cognates, the locus of the emotionality effect should be the phonological form. We consider two possible accounts: (1) the emotionality effect arises from the manner of accessing the emotional concept from the phonological form; (2) the phonological form is associated with episodic emotional memory in the mental lexicon. Regarding the first possibility, following the hierarchical model reviewed above, L2 forms' access to the emotional concept may require the mediation of the corresponding L1 forms, whereas L1 forms can access the emotional concept directly. For Cantonese–Mandarin bilinguals, while the lemma may be linked to both Cantonese and Mandarin

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

phonological forms (see Figure 1), the link with the Cantonese form is probably stronger than the link with the Mandarin form. Thus, accessing the lemma and the higher-level emotional concept would be more efficient from the Cantonese side than from the Mandarin side, which could lead to high automaticity and stronger magnitude in emotional arousal from the Cantonese spoken forms.

The subjective frequency ratings we collected provide some support for the claim of easier lexical and conceptual activation from Cantonese spoken forms. Cantonese pronunciations of the test words were rated as more frequently heard than the Mandarin counterparts, suggesting that the bilinguals are indeed more familiar with the Cantonese forms. However, it is unlikely for ease of activation to account for heightened pupil responses in this study. As the modeling results suggest, subjective frequency *per se* tends to be negatively associated with pupil dilations, with more frequent/familiar forms eliciting weaker pupillary responses, probably due to lower processing effort. This is consistent with previous reports of the effects of lexical frequency on pupil dilation (Kuchinke et al., 2007; Schmidtke, 2014). In other words, the overall effect of ease of activation in the current dataset is weakening – as opposed to strengthening – the pupil response; as a result, ease of activation fails to provide a viable explanation for the observed pattern of heightened emotionality effect in L1 Cantonese.

The second account posits that emotionality is encoded in the phonological forms in memory. The association between emotionality and word pronunciations can be conveniently explained by an exemplar-based model of the lexicon (Johnson, 1997; Pierrehumbert, 2002), where a lexical unit is represented by a cloud of tokens (i.e., exemplars). Each exemplar represents a phonetic form that the individual has encountered previously, including not only the phonetic detail but also the contextual information of the encounter such as the speaker's and the listener's emotional states. Thus, affective words, which tend to occur in emotional context, will have emotional memory associated with their exemplars. In subsequent language processing, the associated emotional memory may be invoked when the affective words are encountered again. For a bilingual speaker, since L2 affective words are less often encountered in naturally-occurring emotional context, the strength of emotional memory associated with an L2 affective word will be much lower compared to its L1 counterpart, thus explaining the differential affective processing phenomenon. Amengual (2012) suggested that cross-linguistic cognates co-inhabit in a cloud in the bilingual lexicon: "Bilinguals may associate two phonologically similar word representations (cognates) in the same ['cloud'], so the word for a particular concept/meaning is influenced by the orthographically, phonologically, and semantically similar representation from the other language" (Amengual, 2012, pp. 527). Our results indicate that even if phonologically similar cognates occupy similar locations in the mental acoustic space, the emotionality information stored in individual exemplars can differ significantly between the two languages. That is to say, cross-linguistic influence within a cloud – if present – is not strong enough to fully equalize the L1-L2 difference in emotional strength.

Obviously, the second account is fully compatible with the idea that emotionality effects are formed through emotional memory in the process of language learning and language use (e.g., Dewaele, 2010; Harris et al., 2006). It also receives support from previous findings of emotionality effects elicited by nonlinguistic auditory stimuli (Blasi et al., 2011; Partala & Surakka, 2003) and enhanced emotionality effects from auditorily presented – compared with visually presented – word stimuli (Harris et al., 2003).

The two accounts are not mutually exclusive. Potential evidence for the first account may come as a result of separating processing effort and efficiency of lexical/conceptual access. The second account could be further tested by examining if the degree of the emotionality effect varies with the fine phonetic detail in the auditory stimuli. Presumably, if the auditory token is phonetically more similar to previously stored exemplars that have high emotionality (e.g., produced with features of emotional speech, such as emotional prosody), the elicited emotionality effect should be stronger. These predictions await confirmation in future research.

L2 proficiency and differential affective processing

It is widely acknowledged that the degree of differential affective processing is smaller if the bilingual is more balanced between L1 and L2. Previous literature suggests that L2 proficiency, early exposure, learning mode (classroom instruction vs. immersion) and language use pattern as potential factors that would determine whether the bilingual would show the emotionality effects in the L2 (Caldwell-Harris, 2014; Degner et al., 2012; Dewaele, 2010; Harris et al., 2006). In the current study, more than half of the participants (34 out of 52) rated their Mandarin proficiency as intermediate, and the rest split between low-proficiency (8) and high-proficiency (10). It should be noted that although this group of bilinguals may lack early childhood exposure to Mandarin in an immersive environment, their L2 experience is overall quite extensive for adult L2 speakers. Even the self-reported low-proficiency group started to learn Mandarin early, between the age of three and seven, and had regular Mandarin lessons throughout primary and secondary schools. Given the linguistic and geographical proximity between Mandarin and Cantonese, it is also safe to assume that the bilinguals in this study have ample exposure to Mandarin in everyday life (especially in recent years) through media and personal contact with Mandarin speakers.

Against the backdrop of overall extensive L2 experience, our results reveal a continuum from self-rated low-proficiency bilinguals, who show the greatest amount of differential affective processing, to mid-proficiency bilinguals, who have a reduced magnitude of the effect, to high-proficiency bilinguals, who show no significant difference in affective processing between L1 and L2. This pattern supports the prediction that the more balanced the bilingual is, the less the differential affective processing effect. Furthermore, the fact that differential affective processing is evidenced in early L2 learners with low-to-mid proficiency levels calls for further research to separate the effects of age of acquisition, learning mode, and proficiency.

Selective effects of differential affective processing

The last question we would like to address is the selectivity of the L1-L2 differences in emotionality effects. In the current study, we find the strongest emotionality effects – as well as the largest L1-L2 differences – with taboo words, while the emotionality effects in emotion-laden words or emotion words are not significant. This pattern is consistent with previous findings of the most pronounced emotionality effects from taboos, swear words, and reprimands. There are at least two mechanisms for taboo words to generate enhanced emotionality effects compared with other

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

word categories. On one hand, taboo words tend to evoke stronger emotional memories; on the other hand, given the social expectation of avoiding taboo words – especially in an academic context – it is possible that taboo words elicit greater surprisal effects in the experiment, which would in turn lead to higher pupil responses. It should be noted, though, that the current study only used taboo words from the standard (i.e., non-slang) Chinese vocabulary, excluding vernacular taboo and swear words that are unique to the spoken variety, which would probably elicit even stronger responses. Thus, our results demonstrate the robustness of the emotionality effects associated with taboo words.

The patterns of differential affective processing observed for taboo words are also in line with previous literature that found larger L1-L2 differences in strongly negative stimuli than in positive/neutral stimuli (Altarriba & Santiago-Rivera, 1994; Bond & Lai, 1986; Dewaele & Pavlenko, 2002; Heredia & Altarriba, 2001; Marian & Kaushanskaya, 2008; Wu & Thierry, 2012). Why should negativity be a modulating factor for differential affective processing? It probably has to do with how humans process negative stimuli. As a crucial feature for survival, we are naturally more alerted to negative stimuli in the environment (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001), a bias that is also reflected in the disproportionately large share of negatively-valenced words in the affective lexicon (Bardeen & Daniel, 2017; Blasi et al., 2011; Cisler & Koster, 2010; Janschewitz, 2008; Lin & Yao, 2016). Disengagement mechanisms such as emotion regulation (see Gratz & Roemer, 2004; Gross, 1998) are available as a way to prevent emotional distress, but such mechanisms could also lead to heightened sensitivity if we need to constantly monitor our attention level to the negative stimuli that are to be avoided (Bardeen & Daniel, 2017; Wenzlaff & Wegner, 2000). Following this line of reasoning, bilinguals may show a lower degree of the negativity bias in the L2 (and consequently greater L1-L2 differences), due to either a general reduction of emotional responses to aversive stimuli in the L2 or a more effective suppression mechanism for the L2 that does not require constant monitoring.

Conclusion

In this paper, we report a pupillometry study that found stronger pupil responses to the L1 (Cantonese) affective words than to the L2 (Mandarin) affective words from Cantonese–Mandarin bilinguals. Our results provide strong support for differential affective processing from a highly restrictive context, i.e., when the tested items are all identical cognates that only differ in pronunciation, suggesting that the emotionality effects in language processing must be at least partially rooted in the phonological forms. This research also extends the work on bilingual affective language processing by including a less-studied L1 and L2 pair.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728922000931
Supplementary Materials

S1. Complete word list used in the study.
S2. Figure of baseline pupil diameter over trials by block by language.
S3. Figure of average task-evoked pupil diameter in the 3000 ms analysis window, separated by block by language.
S4. Summary of `gamm.1a.AR` in the analysis of Language × WordType effects.
S5. Visualization of the effects of IsBlock2, TrialID, and SubjectiveFreq over time in `gamm.1a.AR`.
S6. Summary of `gamm.1b.AR` in the analysis of Language × WordType effects.
S7. Visualization of the partial effects of IsTB and IsC.TB over time in `gamm.1b.AR`.
S8. Summary of `gamm.2a.AR` in the analysis of Language × WordType × MandarinProficiency effects
S9. Summary of `gamm.2b.AR` in the analysis of Language × WordType × MandarinProficiency effects
S10. Visualization of the partial effects of the critical binary variables over time for low-proficiency (top row), mid-proficiency (middle row), and high-proficiency (bottom row) bilinguals in `gamm.2b.AR`.

Acknowledgement. This research project is supported by a research grant from the Hong Kong Polytechnic University (P0001884). We thank Ms. Liu Chang, Ms. Hillarie Tse, and Mr. Eugene Wong for assistance with the experiments.

Competing interests. The authors declare none.

Data Availability Statement. The datasets and the R script for the analysis presented in this paper are publicly accessible from the website https://osf.io/8pgkj/.

References

Altarriba J (2001) Semantics and emotion within and across languages. *J.-M. Dewaele & A. Pavlenko* (*Organizers*), *Linguistic and Conceptual Representations in Bilingual Memory. Colloquium Conducted at the Third International Symposium on Bilingualism, Bristol, England.*

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

Altarriba J and Basnight-Brown DM (2012) The acquisition of concrete, abstract, and emotion words in a second language. *International Journal of Bilingualism* **16**, 446–452. https://doi.org/10.1177/1367006911429511

Altarriba J and Canary T (2004) The influence of emotional arousal on affective priming in monolingual and bilingual speakers. *Journal of Multilingual and Multicultural Development* **25**(June 2014), 37–41. https://doi.org/10.1080/01434630408666531

Altarriba J and Santiago-Rivera AL (1994) Current perspectives on using linguistic and cultural factors in counseling the Hispanic client. *Professional Psychology: Research and Practice* **25**, 388–397. https://doi.org/10.1037/0735-7028.25.4.388

Amengual M (2012) Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition* **15**, 517–530. https://doi.org/10.1017/S1366728911000460

Anooshian L and Hertel P (1994) Emotionality in free speech: Specificity in bilingual memory. *Cognition and Emotion* **8**, 503–514.

Ayçiçegi-Dinn A and Caldwell-Harris CL (2009) Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition* **12**, 291–303. https://doi.org/10.1017/S1366728909990125

Bardeen JR and Daniel TA (2017) An Eye-Tracking Examination of Emotion Regulation, Attentional Bias, and Pupillary Response to Threat Stimuli. *Cognitive Therapy and Research* **0**, 1–14. https://doi.org/10.1007/s10608-017-9860-y

Baumeister RF, Bratslavsky E, Finkenauer C and Vohs KD (2001) Bad Is Stronger Than Good. *Review of General Psychology* **5**, 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bayer M, Sommer W and Schacht A (2011) Emotional words impact the mind but not the body: Evidence from pupillary responses. *Psychophysiology* **48**, 1554–1562. https://doi.org/10.1111/j.1469-8986.2011.01219.x

Blasi A, Mercure E, Lloyd-Fox S, Thomson A, Brammer M, Sauter D, … Murphy DGM (2011) Early specialization for voice and emotion processing in the infant brain. *Current Biology* **21**, 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009

Bond MH and Lai T-M (1986) Embarrassment and code-switching into a second language. *The Journal of Social Psychology* **126**, 179–186.

Bowers JS and Pleydell-Pearce CW (2011) Swearing, euphemisms, and linguistic relativity. *PLoS ONE* **6**. https://doi.org/10.1371/journal.pone.0022341

Bradley MM, Miccoli L, Escrig MA and Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Brown E and Harper D (2009) Phonological evidence of crosslinguistic exemplar connections. *Studies in Hispanic and Lusophone Linguistics* **2**, 257–274. https://doi.org/10.1515/shll-2009-1052

Cai ZG, Pickering MJ, Yan H and Branigan HP (2011) Lexical and syntactic representations in closely related languages: Evidence from Cantonese–Mandarin bilinguals. *Journal of Memory and Language* **65**, 431–445. https://doi.org/10.1016/j.jml.2011.05.003

Caldwell-Harris CL (2014) Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology* **5**, 1–4. https://doi.org/10.3389/fpsyg.2014.01055

Caldwell-Harris CL (2015) Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science* **24**, 214–219. https://doi.org/10.1177/0963721414566268

Caldwell-Harris CL, Tong J, Lung W and Poo S (2011) Physiological reactivity to emotional phrases in Mandarin-English bilinguals. *International Journal of Bilingualism* **15**, 329–352. https://doi.org/10.1177/1367006910379262

Chen P, Lin J, Chen B, Lu C and Guo T (2015) Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese-English bilinguals. *Cortex* **71**, 34–48. https://doi.org/10.1016/j.cortex.2015.06.002

Christensen RHB (2019) *ordinal - Regression Models for Ordinal Data*. Retrieved from https://cran.r-project.org/package=ordinal

Cisler JM and Koster EHW (2010) Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review* **30**, 203–216. https://doi.org/10.1016/j.cpr.2009.11.003

Cochrane RM (1980) The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language. *Journal OfMultilingual and Multicultural Development* **1**, 331–360.

Costa A, Foucart A, Corey JD, Keysar B, Hayakawa S, Botella J and Aparici M (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

Costa A, Foucart A, Hayakawa S, Aparici M, Apesteguia J, Heafner J and Keysar B (2014) Your morals depend on language. *PLoS ONE* **9**. https://doi.org/10.1371/journal.pone.0094842

Costa A, Santesteban M and Caño A (2005) On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language* **94**, 94–103. https://doi.org/10.1016/j.bandl.2004.12.002

Degner J, Doycheva C and Wentura D (2012) It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition* **15**(Special Issue 01), 181–189. https://doi.org/10.1017/S1366728911000095

Dewaele J-M (2004) The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development* **25**, 204–222. https://doi.org/10.1080/01434630408666529

Dewaele J-M (2008) The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics* **40**, 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002

Dewaele J-M (2010) *Emotions in Multiple Languages*. https://doi.org/10.1057/9780230289505

Dewaele J-M and Pavlenko A (2002) Emotion vocabulary in interlanguage. *Language Learning* **52**, 263–322. https://doi.org/10.1111/0023-8333.00185

Dijkstra T, Hilberink-Schulpen B and Van Heuven WJB (2010) Repetition and masked form priming within and between languages using word and nonword neighbors. *Bilingualism: Language and Cognition* **13**, 341–357. https://doi.org/10.1017/S1366728909990575

Eilola TM and Havelka J (2011) Behavioural and physiological responses to the emotional and taboo stroop tasks in native and non-native speakers of english. *International Journal of Bilingualism* **15**, 353–369. https://doi.org/10.1177/1367006910379263

Eilola TM, Havelka J and Sharma D (2007) Emotional activation in the first and second language. *Cognition and Emotion* **21**, 1064–1076. https://doi.org/10.1080/02699930601054109

Ekman P (1992) An argument for basic emotions. *Cognition and Emotion* **6**, 169–200.

Fan L, Xu Q, Wang X, Zhang F, Yang Y and Liu X (2016) Neural correlates of task-irrelevant first and second language emotion words - Evidence from the emotional face-word stroop task. *Frontiers in Psychology* **7**, 1–13. https://doi.org/10.3389/fpsyg.2016.01672

Ferré P, García T, Fraga I, Sánchez-Casas R and Molero M (2010) Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition & Emotion* **24**, 760–785. https://doi.org/10.1080/02699930902985779

Ferré P, Sánchez-Casas R and Fraga I (2013) Memory for emotional words in the first and the second language: Effects of the encoding task. *Bilingualism: Language and Cognition* **16**, 495–507. https://doi.org/10.1017/S1366728912000314

Flege J, Frieda E, Walley A and Randazza L (1998) Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition* **20**, 155–187.

Flege J and Munro MJ (1994) The word unit in second language speech production and perception. *Studies in Second Language Acquisition* **16**, 381–411. https://doi.org/10.1017/S0272263100013437

Flege J, Takagi N and Mann V (1995) Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech* **38**, 25–55.

Gagl B, Hawelka S and Hutzler F (2011) Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods* **43**, 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Gao S, Zika O, Rogers RD and Thierry G (2015) Second language feedback abolishes the "hot hand" effect during even-probability gambling. *The Journal of Neuroscience* **35**, 5983–5989. https://doi.org/10.1523/jneurosci.3622-14.2015

Garrido MV and Prada M (2018) Comparing the valence, emotionality and subjective familiarity of words in a first and a second language. *International Journal of Bilingual Education and Bilingualism* **0**, 1–17. https://doi.org/10.1080/13670050.2018.1456514

Goldrick M, Runnqvist E and Costa A (2014) Language switching makes pronunciation less nativelike. *Psychological Science* **25**, 1031–1036. https://doi.org/10.1177/0956797613520014

Gratz KL and Roemer L (2004) Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment* **26**, 41–54. https://doi.org/10.1023/B:JOBA.0000007455.08539.94

Grosjean F (1998) Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* **1**, 131–149.

Gross JJ (1998) The emerging field of emotion regulation: An integrative review. *Review of General Psychology* **2**, 271–299.

Harris CL, Ayçiçegi A and Gleason JB (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

Harris CL, Gleason JB and Ayçiçeği A (2006) When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In *Bilingual Minds*. pp. 257–283 https://doi.org/10.1.1.69.5830

Hartmann M and Fischer MH (2014) Pupillometry: The eyes shed fresh light on the mind. *Current Biology* **24**, R281–R282. https://doi.org/10.1016/j.cub.2014.02.028

Heredia RR and Altarriba J (2001) Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science* **10**, 164–168. https://doi.org/10.1111/1467-8721.00140

Hess EH (1965) Attitude and pupil size. *Scientific American* **212**, 46–54.

Hsu CT, Jacobs AM and Conrad M (2015) Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex* **63**, 282–295. https://doi.org/10.1016/j.cortex.2014.09.002

Huang Y and Snedeker J (2020) Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition* **200**(December 2017), 104251. https://doi.org/10.1016/j.cognition.2020.104251

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

Iacozza S, Costa A and Duñabeitia JA (2017) What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PLoS ONE* **12**, 1–10. https://doi.org/10.1371/journal.pone.0186027

Janisse MP (1974) Pupil size, affect and exposure frequency. *Social Behavior and Personality* **2**, 125–146.

Janschewitz K (2008) Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods* **40**, 1065–1074. https://doi.org/10.3758/BRM.40.4.1065

John CH (1988) Emotionality ratings and free-association norms of 240 emotional and non-emotional words. *Cognition and Emotion* **2**, 49–70.

Johnson K (1997) Speech perception without speaker normalization. In Johnson K and Mullennix JW (eds), *Talker Variability in Speech Processing*. San Diego, CA: Academic Press, pp. 145–166.

Jończyk R, Boutonnet B, Musiał K, Hoemann K and Thierry G (2016) The bilingual brain turns a blind eye to negative statements in the second language. *Cognitive, Affective and Behavioral Neuroscience* **16**, 527–540. https://doi.org/10.3758/s13415-016-0411-x

Kazanas SA and Altarriba J (2015) The automatic activation of emotion and emotion-laden words: Evidence from a masked and unmasked priming paradigm. *American Journal of Psychology* **128**, 323–336. https://doi.org/10.5406/amerjpsyc.128.3.0323

Kazanas SA and Altarriba J (2016) Emotion word processing: Effects of word type and valence in Spanish–English bilinguals. *Journal of Psycholinguistic Research* **45**, 395–406. https://doi.org/10.1007/s10936-015-9357-3

Kensinger EA and Corkin S (2004) Two routes to emotional memory: distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3310–3315. https://doi.org/10.1073/pnas.0306408101

Keysar B, Hayakawa SL and An SG (2012) The Foreign-Language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science* **23**, 661–668. https://doi.org/10.1177/0956797611432178

Kissler J and Strehlow J (2017) Something always sticks? how emotional language modulates neural processes involved in face encoding and recognition memory. *Poznan Studies in Contemporary Linguistics* **53**, 63–93. https://doi.org/10.1515/psicl-2017-0004

Knapen T, De Gee JW, Brascamp J, Nuiten S, Hoppenbrouwers S and Theeuwes J (2016) Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS ONE* **11**, 1–13. https://doi.org/10.1371/journal.pone.0155574

Koelewijn T, Zekveld AA, Festen JM and Kramer SE (2012) Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing* **33**, 291–300. https://doi.org/10.1097/AUD.0b013e3182310019

Kroll JF and de Groot AM (1997) Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In de Groot AMB and Kroll JF (eds), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Lawrence Erlbaum Associates, Publishers.

Kroll JF and Stewart E (1994) Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, Vol. **33**, pp. 149–174. https://doi.org/10.1006/jmla.1994.1008

Ksiazek P, Wendt D, Alickovic E and Lunne T (2018) Analysis of the individual listening effort reflected by the pupillary responses during speech perception in noise. *Poster at the* 10*th Speech in Noise Workshop*, (2013), 2017.

Kuchinke L, Võ MLH, Hofmann M and Jacobs AM (2007) Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology* **65**, 132–140. https://doi.org/10.1016/j.ijpsycho.2007.04.004

Lalor E and Kirsner K (2000) Cross-lingual transfer effects between English and Italian cognates and noncognates. *International Journal of Bilingualism* **4**, 385–398. https://doi.org/10.1177/13670069000040030501

Lin J and Yao Y (2016) Encoding emotion in Chinese: a database of Chinese emotion words with information of emotion type, intensity, and valence. *Lingua Sinica* **2**, 6. https://doi.org/10.1186/s40655-016-0015-y

Maddock RJ, Garrett AS and Buonocore MH (2003) Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping* **18**, 30–41. https://doi.org/10.1002/hbm.10075

Manning SK and Melchiori MP (1974) Words that upset urban college students: Measured with gsrs and rating scales. *Journal of Social Psychology* **94**, 305–306. https://doi.org/10.1080/00224545.1974.9923225

Maratos EJ, Allan K and Rugg MD (2000) Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia* **38**, 1452–1465. https://doi.org/10.1016/S0028-3932(00)00061-0

Marian V and Kaushanskaya M (2008) Words, feelings, and bilingualism: Cross-linguistic differences in emotionality of autobiographical memories. *Mental Lexicon* **3**, 72–90. https://doi.org/doi:10.1075/ml.3.1.06mar

Mathôt S, Fabius J, Van Heusden E and Van Der Stigchel S (2018) Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods* **50**, 94–106. https://doi.org/10.3758/s13428-017-1007-2

McGinley MJ, David SV and McCormick DA (2015) Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* **87**, 179–192. https://doi.org/10.1016/j.neuron.2015.05.038

Milanović I (2019) *English as a foreign language: How emotional can it be?* University of Zagreb.

Min CS and Schirmer A (2011) Perceiving verbal and vocal emotions in a second language. *Cognition & Emotion* **25**, 1376–1392. https://doi.org/10.1080/02699931.2010.544865

Mortier A (2013) *Processing neutral words, negative words and taboo words: An ERP study*. University Gent.

Oatley K and Johnson-Laird PN (1987) Towards a cognitive theory of emotions. *Cognition and Emotion* **1**, 29–50.

Oganian Y, Korn CW and Heekeren HR (2016) Language switching-but not foreign language use per se-reduces the framing effect. *Journal of Experimental Psychology: Learning Memory and Cognition* **42**, 140–148. https://doi.org/10.1037/xlm0000161

Ong E, Hussain S, Chow Y and Thompson C (2017) Variations in Bilingual Processing of Positive and Negative Information. *The* 6*th Annual International Conference on Cognitive and Behavioral Psychology*, 36–42. https://doi.org/10.5176/2251-1865_cbp17.14

Opitz B and Degner J (2012) Emotionality in a second language: It's a matter of time. *Neuropsychologia* **50**, 1961–1967. https://doi.org/10.1016/j.neuropsychologia.2012.04.021

Partala T and Surakka V (2003) Pupil size variation as an indication of affective processing. *International Journal of Human Computer Studies* **59**, 185–198. https://doi.org/10.1016/S1071-5819(03)00017-X

Pavlenko A (1999) New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition* **2**, 209–230. https://doi.org/10.1017/S1366728999000322

Pavlenko A (2008) Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition* **11**, 147–164. https://doi.org/10.1017/S1366728908003283

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

Pavlenko A (2017) Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current Opinion in Psychology* **17**, 74–78. https://doi.org/10.1016/j.copsyc.2017.06.005

Peeters D, Dijkstra T and Grainger J (2013) The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language* **68**, 315–332. https://doi.org/10.1016/j.jml.2012.12.003

Pierrehumbert JB (2002) Word-specific phonetics. In Gussenhoven C and Warner N (eds), *Laboratory Phonology {VII}* (Vol. 7), Berlin, New York: Mouton de Gruyter, pp. 101–140.

Plutchik R (1980) *Emotions: A psychoevolutionary synthesis*. New York, NY: Harper and Row.

Ponari M, Rodríguez-Cuadrado S, Vinson D, Fox N, Costa A and Vigliocco G (2015) Processing advantage for emotional words in bilingual speakers. *Emotion* **15**, 644–652. https://doi.org/10.1037/emo0000061

Recio G, Conrad M, Hansen LB and Jacobs AM (2014) On pleasure and thrill: The interplay between arousal and valence during visual word recognition. *Brain and Language* **134**, 34–43. https://doi.org/10.1016/j.bandl.2014.03.009

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press

Robinson CJ and Altarriba J (2018) The interrelationship between emotion, cognition, and bilingualism. *Yearbook of the Poznan Linguistic Meeting* **1**, 103–117. https://doi.org/10.1515/yplm-2015-0006

Robison MK and Unsworth N (2019) Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics* **81**, 407–419.

Rozin P and Royzman EB (2001) Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* **5**, 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Schacht AK, Dimigen O and Sommer W (2010) Emotions in cognitive conflicts are not aversive but are task specific. *Cognitive, Affective and Behavioral Neuroscience* **10**, 349–356. https://doi.org/10.3758/CABN.10.3.349

Schmidtke J (2014) Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology* **5**, 1–16. https://doi.org/10.3389/fpsyg.2014.00137

Segalowitz N, Trofimovich P, Gatbonton E and Sokolovskaya A (2008) Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon* **3**, 47–71. https://doi.org/10.1075/ml.3.1.05seg

Simcox T, Pilotti M, Mahamane S and Romero E (2012) Does the language in which aversive stimuli are presented affect their processing? *International Journal of Bilingualism* **16**, 419–427. https://doi.org/10.1177/1367006911425821

Simonet M and Amengual M (2020) Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. *International Journal of Bilingualism* **24**, 208–221. https://doi.org/10.1177/1367006919826388

Sóskuthy M (2017) *Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction*. Retrieved from http://arxiv.org/abs/1703.05339

Sóskuthy M (2021) Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* **84**. https://doi.org/10.1016/j.wocn.2020.101017

SR Research (2011) *SR Research Experiment Builder [Computer software]*. Mississauga, Ontario, Canada: SR Research Ltd.

Stankovic M, Biedermann B and Hamamura T (2022) Not all bilinguals are the same: A meta-analysis of the moral foreign language effect. *Brain and Language* **227**, 105082. https://doi.org/10.1016/j.bandl.2022.105082

Strauss G and Allen D (2008) Emotional intensity and categorisation ratings for emotional and nonemotional words. *Cognition & Emotion* **22**, 114–133. https://doi.org/10.1080/02699930701319154

Sutton TM, Altarriba J, Gianico JL and Basnight-Brown DM (2007) The automatic access of emotion: Emotional Stroop effects in Spanish–English bilingual speakers. *Cognition and Emotion* **21**, 1077–1090. https://doi.org/10.1080/02699930601054133

Talmi D and Moscovitch M (2004) Can semantic relatedness explain the enhancement of memory for emotional words? *Memory and Cognition* **32**, 742–751. https://doi.org/10.3758/BF03195864

Tamási K, McKean C, Gafos A, Fritzsche T and Höhle B (2017) Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology* **153**, 140–148. https://doi.org/10.1016/j.jecp.2016.07.014

Team RC (2019) *R: A language and environment for statistical computing*. Retrieved from https://www.r-project.org/

Tryon WW (1975) Pupillometry: A survey of sources of variation. *Psychophysiology* **12**, 90–93.

Turner JH (2000) *On the origins of human emotions: A sociological inquiry into the evolution of human affect*. Sanford, CA: Stanford University Press.

van Rij J, Hendriks P, van Rijn H, Baayen RH and Wood SN (2019) Analyzing the time course of p upillometric data. *Trends in Hearing* **23**, 1–22. https://doi.org/10.1177/2331216519832483

van Rij J, Wieling M, Baayen R and van Rijn H (2017) Itsadug: Interpreting time series and autocorrelated data using GAMMs. *R Package Version 2. 4.*

Võ MLH, Jacobs AM, Kuchinke L, Hofmann M, Conrad M, Schacht A and Hutzler F (2008) The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology* **45**, 130–140. https://doi.org/10.1111/j.1469-8986.2007.00606.x

Wagner AE, Nagels L, Toffanin P, Opie JM and Başkent D (2019) Individual variations in effort: Assessing pupillometry for the hearing impaired. *Trends in Hearing* **23**, 1–18. https://doi.org/10.1177/2331216519845596

Wenzlaff RM and Wegner DM (2000) Thought suppression. *Annual Review of Psychology* **51**, 59–91.

Wieling M (2018) Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* **70**, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Winn MB, Wendt D, Koelewijn T and Kuchinsky SE (2018) Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing* **22**, 1–32. https://doi.org/10.1177/2331216518800869

Winskel H (2013) The emotional Stroop task and emotionality rating of negative and neutral words in late Thai-English bilinguals. *International Journal of Psychology* **48**, 1090–1098.

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 3–36.

Wood SN (2017) Generalized Additive Models: An Introduction with R. In *Generalized Additive Models* (2nd editio). https://doi.org/10.1201/9781315370279-4

Wu YJ and Thierry G (2012) How reading in a second language protects your heart. *Journal of Neuroscience* **32**, 6485–6489. https://doi.org/10.1523/jneurosci.6119-11.2012

Yao Y and Chang CB (2016) On the cognitive basis of contact-induced sound change: vowel merger reversal in Shanghainese. *Language* **92**, 433–467. https://doi.org/10.1017/CBO9781107415324.004

https://doi.org/10.1017/S1366728922000931 Published online by Cambridge University Press