ARTICLE

OPEN

Check for updates

Why do Mandarin speakers code-switch? A case study of conversational code-switching in China

Xinyi Zhong[1], Lay Hoon Ang [1✉] & Sharon Sharmini[1]

Multilingualism is becoming increasingly common worldwide, with multilingual studies becoming more popular and accelerating interpersonal language contacts; however, it remains controversial and ambiguous in China. The current study aims to explore code-switching (CS) patterns and factors among Mandarin-dominated mainland Chinese speakers and investigate multilingual development in China through CS, as it is a multilingual practice. Conversational analysis was conducted to analyse daily interactions in short videos posted online, categorise language varieties and frequencies, identify CS patterns, and examine the factors influencing CS patterns. The results revealed that foreign languages and dialects other than Mandarin were spoken by Mandarin-dominated speakers. Three of four CS patterns were also pinpointed regarding language practices in China. Insertion was the most predominant pattern, followed by backflagging and alternation. The CS patterns employed by Mandarin-dominated mainland Chinese speakers were influenced primarily by participant-related factors, followed by linguistic-related factors, sociopsychological factors, convenience, and situational factors. The prevalence of the insertional CS pattern and the influence of personal factors suggest that China's Mandarin-speaking community is in an early phase of multilingual development.

[1] Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia. ✉email: hlang@upm.edu.my

Introduction

Global interactions have accelerated in recent decades due to economic and technological globalisation. Multilingualism, which emerges from global interactions, is a worldwide phenomenon (Wardhaugh and Fuller 2015). It is a social situation involving groups or communities communicating in more than one language with varying proficiency levels (Maher, 2017). The debate on whether China is a multilingual society remains unresolved despite the increasing trend of multilingualism. Nevertheless, China is one of the countries with the richest linguistic diversity, with 306 established languages distributed across 56 nationalities (Eberhard et al. 2021). The Chinese populace speaks a variety of languages and dialects, with Mandarin being the most widely spoken and serving as the official language of China. Mandarin is the medium of school instruction and is generally employed in official settings, such as the government, education, and mass media. Simultaneously, other regional dialects, including Cantonese, Wu, Min, and Hakka, are spoken by individuals from each respective region, primarily with family and friends in informal settings. Several minority languages (Tibetan, Uyghur, Mongolian, and Korean) are spoken in China in addition to foreign languages, including English, Japanese, and Russian. The languages comprise unique writing systems and cultural traditions that are indispensable to China's rich linguistic and cultural landscapes.

The Chinese exhibit particularly diverse and constantly evolving language practices that reflect a rich linguistic heritage and ongoing integration into the global community. Moreover, employing multiple languages may lead to the development of multilingual practices, such as diglossia, code-switching (CS), code-mixing, borrowing, and translanguaging. The current study seeks to explore a common linguistic phenomenon, namely, CS, to determine whether China is a multilingual country (Carstens and Ang 2019). CS refers to the mixture of more than one code within a discourse or communication (Myers-Scotton 2002). It is regarded as the most creative communicative strategy in a multilingual community (Hoffmann, 2014; Li 2011).

However, the existence of CS in China remains a subject of debate. Zhang (2005) defined CS as bidialectalism instead of multilingualism upon analysing Mandarin–Cantonese CS practices across radio conversations in Shenzhen. Li (2006) also described China's language policy and goal as bidialectalism or dialect bilingualism. Furthermore, previous CS studies in China have focused more on CS practices in Hong Kong, Macao, and Taiwan (Li and Elly 2002; Chan 2018) than in mainland China. However, the language practices of Chinese mainlanders have been neglected, which also manifests rich language mixings among foreign languages in addition to Mandarin and the corresponding dialects. Several studies conducted in mainland China have focused only on bilingual teachers and students or specific regions. Hence, further investigation into CS practices in mainland China is imperative to understand multilingualism in China. The present study aims to study Mandarin-dominated CS practices by considering the controversies and complexity of multilingualism in China given the dominance of Mandarin. Muysken's (2000, 2013) typology and Ritchie and Bhatia's (2013) model were applied to explore the patterns and factors of CS, respectively, to examine multilingualism in China.

Literature review

CS is defined as instances in which "lexical items and grammatical features from two languages appear in one sentence" (Muysken, 2000: 1) and has been studied from various syntactic, sociolinguistic, psycholinguistic, and interdisciplinary perspectives (Zhong et al. 2023). This section reviews CS studies from syntactic and sociolinguistic perspectives to explore relevant CS patterns and factors. Identifying CS patterns is generally the initial step in determining CS practices in a specific region. Poplack's (1980) trichotomy, including intra-sentential, inter-sentential, and tag switching, is frequently employed to analyse CS within and between sentences, clauses, and contextualisation cues, respectively. Poplack's (1980) trichotomy also focuses on CS placement inside or outside a sentence rather than the grammatical structures emphasised by Muysken (2000, 2013).

Muysken's (2000, 2013) typology, which divides CS patterns into the four types of insertion, alternation, congruent lexicalisation, and backflagging, is another common theoretical framework. Insertion refers to the situation in which Language B is inserted into a sentence of Language A at the phonological, morphological, lexical, phrasal, or clausal level. Languages A and B adhere to their respective grammatical structures within a sentence, while the entire discourse remains dominated by Language A in alternation. Congruent lexicalisation occurs when the dominant language is ambiguous and other languages with similar grammatical structures are evenly mingled with one another. Backflagging, which was proposed by Muysken in 2013 upon defining the first three CS patterns in 2000, is the reverse insertion with Language A inserted into a sentence following the grammatical structure of Language B in a Language A-dominated discourse.

As verified by multiple scholars, Muysken's (2000, 2013) typology could be adapted to empirical studies and feasible in specific contexts. However, the typology has received only partial support from certain scholars. Albirini and Chakrani (2017) only provided evidence of insertion in Arabic and English CS practices. Vaughan (2021) roughly divided CS patterns into inter-sentential and intra-sentential, which includes both insertional and alternational switching.

In contrast, Stell (2010) corroborated that Muysken's (2000) three CS patterns of insertion, alternation and congruent lexicalisation were applicable in the South African context or in Afrikaans–English CS practices. Insertion is the most common pattern in Luxembourg, followed by alternation, while congruent lexicalisation is less frequent (Stell and Couto, 2012). Alternation is the most frequent practice in the United States of America (USA), followed by insertion and congruent lexicalisation (Mad'arová, 2018). Wu et al. (2022) found that insertion, inter-sentential, and backflagging are more dominant CS patterns among Singaporean preschoolers than alternation. Yahiaoui et al. (2021) confirmed Muysken's (2000, 2013) typology in the Lebanese context and identified the syntactic elements of each pattern. Under the insertional pattern, the five syntactic elements of interjection, noun, proper noun, adjective, and politeness sign were discovered.

While previous researchers have partially or completely proven the feasibility of applying Muysken's (2000, 2013) typology to identify CS patterns in various contexts, others have challenged and revised it. Demirçay and Backus (2014) reassessed the typology of CS patterns in the Netherlands. Ambiguity was revealed between insertion and alternation, in which certain examples were difficult to categorise into two patterns. Lipski (2014) examined three CS patterns in the American context and expanded the pattern of congruent lexicalisation to include 'ragged' and possibly involuntary code-mixing among semi-fluent bilinguals. Tramutoli (2021) posited that a mid-step should exist between insertion and alternation. The current study appraised Muysken's (2000, 2013) typology and determined its applicability for identifying CS patterns in China.

The factors influencing CS practices are the central issue among sociolinguists and are generally explored from micro and

macro perspectives. The micro aspect concerns pragmatic factors, such as filling lexical needs and quoting and qualifying a message (Al-Daher, 2021), or pedagogical factors, such as managing classes, establishing rapport, and accessing curricula (May and Aziz 2020). The macro aspect primarily explores factors influencing all speakers in society, including language exposure, proficiency, features, development, and individual differences (Smolak et al. 2020). Fishman (1965) pioneered the idea that language choice is not an arbitrary behaviour but is influenced by the situation, participant, and topic. Similarly, Gumperz (1982) divided CS into situational and metaphorical, wherein CS is triggered by situational and non-situational factors, respectively. Myers-Scotton (1993) investigated social motivations for CS by combining social and personal factors based on the decennial African fieldwork. Myers-Scotton (1993) classified four language choices based on the markedness model, in which CS is either an unmarked choice, a marked choice, a sequential unmarked choice, or an explanatory choice. The first two choices are similar to Gumperz's (1982) situational and metaphorical CS, whereas the latter two choices refer to language choices based on negotiation, the setting, or the topic. Hoffmann (2014), Holmes (2013), and Malik (1994) also evaluated the micro-pragmatic functions of CS.

Several models explain CS factors from both the micro and macro perspectives. Ritchie and Bhatia's (2013) model is a globally popular framework (Fakeye 2012; Dewaele and Zeckel 2016; Prirol and Masruddin 2019). The details are as follows:

1. Participants' social roles and relationships or factors related to the speaker include characteristics such as age, gender, religion, social class, identity, and social distance.
2. Situational factors or factors related to the situation encompass the formality, setting, topic, and social group.
3. Message-intrinsic considerations or factors within the conversation refer to clarification, contrast, hedging, idioms and deep-rooted cultural wisdom, interjection, language trigger, message qualification, paraphrase, quotations, repetition, and topic–comment or relative clauses.
4. Sociopsychological factors include individual and social attitudes, language dominance, linguistic security, group membership, neutrality, and speech accommodation.

The aforementioned factors explicate CS practices, which are context-driven and performed distinctively. Scholars have also tended to search for universal models explaining CS practices in all contexts upon reviewing the existing CS models. However, context-based factors require further exploration. Previous models inspired this study to explore how both micro- and macro-level factors could influence CS practices. Ritchie and Bhatia's (2013) model and a detailed context-oriented explanation were incorporated into this study to assess the factors influencing CS practices in China.

Methods

The current qualitative case study explored CS patterns and factors by analysing CS practices in 16 short videos. Sixteen Chinese participants purposively selected the videos via online platforms. The participants were recruited to collect the data on behalf of the researchers.

Participants who fulfilled the inclusion criteria (see Table 1) were recruited to collect short videos that could represent the speaking styles of mainland Chinese speakers. Eligible participants had to be Mandarin-dominated mainland Chinese speakers, which corresponds to the current research objective of appraising Mandarin-dominated CS practices in mainland China. Mandarin-dominated mainland Chinese speakers are individuals

Table 1 Participant inclusion and exclusion criteria.

Inclusion criteria	Mandarin-dominated mainland Chinese speakers
Exclusion criteria	Non-Mandarin-dominated mainland Chinese speakers

Table 2 Demographic information of the recruited participants.

	N	Gender		Age	
		Male	Female	Range	Mean
Participants	16	5	11	22–32	24.75

with Chinese ethnicity and nationality who self-reported employing Mandarin as their dominant language. Non-Mandarin-dominated speakers were excluded. A total of 16 voluntary participants (five males and 11 females) were recruited one day after the recruitment information was posted on WeChat (social networking software) on September 10, 2021.

Table 2 briefly presents the demographic information of the recruited participants. The participants were between 22 and 32 years old, with a mean age of 24.75 years. None had speaking or hearing impairments.

All 16 participants were requested to select short videos with Mandarin-dominated CS practices by adhering to the inclusion and exclusion criteria (see Table 3). Essentially, the included videos were to contain Mandarin-dominated CS practices. Videos that were dominated by English, dialects, or other languages, or without CS practices were to be excluded. Second, the videos were to include two or more individuals spontaneously communicating with each other, as spontaneous conversations can reveal a natural way of interaction. The criteria excluded videos lacking communication or featuring communication via a prescheduled method, such as monologues, speeches, films, television series, and advertisements. Furthermore, the video duration was to be five to 30 mins. Shorter or longer videos were excluded because they might be too short to effectively demonstrate language practices or too long with multiple repetitions of language practices. Participants were also to collect short videos that could optimally represent the speaking styles of mainland Chinese speakers rather than videos with obscure expressions or deliberate CS practices. Short videos with prankish, vulgar, or sensitive topics were excluded.

Table 4 presents the short video collection procedures used by the participants in China. Each participant was asked to select three short videos by strictly adhering to the inclusion and exclusion criteria. No limit was imposed on the choice of online video platforms, which allowed participants to freely search for videos. The researchers determined whether the selected videos fulfilled the inclusion and exclusion criteria. Participants whose selections did not fulfil the criteria were required to make new selections until every participant selected three short videos that fulfilled the criteria. Subsequently, all 48 short videos were compiled. Each participant voted for the top 16 videos that most optimally represented the speaking styles of mainland Chinese speakers. Each participant possessed 16 votes. If more than 16 videos were selected due to some having identical numbers of votes, the participants voted again for the last several videos. The voting procedure continued until the top 16 videos were finalised. Basic information on the 16 selected short videos is presented in Table 5.

The selected videos were posted on the online platforms Bilibili and Sina Weibo, and the video duration ranged between 7m18s and 27m2s, with a mean duration of 14m9s. The topics of the selected 16 videos include beauty and makeup, casual conversation, digital products, food, health care, marriage, and the workplace. The 16 selected short videos were transcribed and translated for further analysis. Different language varieties were identified and coded before the corresponding frequencies were calculated. The CS patterns were examined adhering to Muysken's (2000, 2013) typology, which divides CS patterns into insertion, alternation, congruent lexicalisation, and backflagging. The factors influencing the CS patterns were explored through Ritchie and Bhatia's (2013) model. Four types of factors were investigated: participants' social roles and relationships, situational factors, message-intrinsic considerations, and socio-psychological factors. The co-raters were advised to code via ATLAS.ti, a computer-assisted qualitative data analysis software. Two different academic raters conducted the coding independently. Cohen's Kappa was utilised to test the inter-rater reliability between the coding made by the two independent raters. The inter-rater reliability was 0.89, indicating an agreement between the two raters (Landis and Koch 1977). All disagreements between the raters were settled through a consensus.

Results and discussion
This study identified the language varieties of mainland Chinese speakers, which are generally classified as Mandarin or non-Mandarin, as a prerequisite for identifying CS patterns. The

frequencies of the two varieties are presented in Table 6. Mandarin was employed in most situations in the videos, with a 98.743% frequency; non-Mandarin language varieties were limited, with a frequency of only 1.257%. The Mandarin-dominated language practices may result from language policies in recent decades, with Mandarin serving as the official language and other language varieties receiving considerable support from the Chinese government.

Mandarin, or Putonghua, is the dominant language in China and was employed as a tool to unify and modernise the nation following the continuous defeats in the 1840s. Formally adopted as the official standard language in 1956, it has been promoted since the late 1950s. Specific laws have recently been implemented to establish the predominance of the language. For example, the "Law of the People's Republic of China on the Standard Spoken and Written Chinese Language" and China's Constitution were enacted to enhance the status of Mandarin (Adamson and Feng 2021). The Chinese government has also continuously strived to popularise Mandarin. Promoting Mandarin is deemed a constitutional duty and an effective approach to developing a strong sense of community. Currently, Mandarin is the national language and instruction medium in China (Zhang and Cai 2020). Its penetration rate had reached 80.72% by 2020 (Yu 2021).

In addition to Mandarin, mainland Chinese speakers use other language varieties, including foreign languages and dialects (see Table 7). English was revealed as the most frequently employed foreign language, followed by Japanese and French. Language practices were established under the influence of foreign language policies launched by the Ministry of Education in China, which have stipulated English as the main foreign language since 1987 (Zhang et al. 2020). Since amendments to the school curriculum in 2001, English has been strongly promoted in education as one of the three main subjects in addition to Mandarin and mathe-

Table 3 Short video inclusion and exclusion criteria.

Inclusion Criteria	1. Videos with Mandarin-dominated CS practices. 2. Two or more individuals spontaneously communicate in the videos. 3. Video duration is between 5 and 30 mins. 4. Videos optimally represent the speaking styles of mainland Chinese speakers.
Exclusion Criteria	1. Videos without Mandarin-dominated CS practices. 2. Monologues, speeches, films, television series, and advertisements. 3. Video duration is under 5 or over 30 mins. 4. Videos with obscure or deliberate CS expressions. 5. Videos on prankish, vulgar or sensitive topics.

Table 4 Procedures for collecting short videos for Chinese speakers.

Data	Data collection procedures
Short videos	1. Each participant selects three short videos. 2. Researchers check the videos selected by each participant. 3. All 48 short videos are collected. 4. Each participant votes for the top 16 short videos. 5. The top 16 short videos are finalised.

Table 6 Frequencies of Mandarin and Non-Mandarin words.

	Total number of words	Percentage/%
Mandarin	63,294	98.743
Non-Mandarin	806	1.257
Total	64,100	100

Table 7 Detailed frequencies of Non-Mandarin words.

	Total words	Percentage/%
Foreign languages	756	93.797
English	750	93.052
Japanese	5	0.621
French	1	0.124
Dialects	50	6.203
Beijing dialect	24	2.978
Northeastern dialect	6	0.744
Sichuan dialect	20	2.481
Total CS words	806	100

Table 5 Basic information on the 16 selected short videos.

	N	Video duration			Online platform	Topic
		Range	**Total**	**Mean**		
Short videos	16	7m18s–27m2s	3h54m15s	14m9s	Bilibili and Sina Weibo	Beauty and makeup, casual conversation, digital products, food, health care, marriage, workplace.

matics starting in primary school. Therefore, the younger generation demonstrates a certain proficiency in speaking English. The inserted Japanese and French utterances did not contain content words but only several idioms, proper nouns, or brand names, none of which required high ability.

Table 7 illustrates that the Beijing, Northeastern and Sichuan dialects are employed by Chinese people with similar frequencies in China. The three dialects were used separately in four of the 16 videos by four speakers from the three different regions. Other regional dialects were also equivalently spoken by Mandarin speakers (Francis, 2016), as dialects currently receive considerable support from the People's Republic of China (PRC) government. Bidnoshyia and Dyka (2022) reported that the prestige of a language is influenced by language policies. The Chinese language policy towards dialects was implemented for three decades after the PRC's establishment. The plan was for regional dialects to be gradually replaced by Mandarin in every domain from the 1950s to the 1980s. However, the relationship between Mandarin and dialects was reconsidered from the 1980s to the early 1990s. Dialect revival emphasises that dialects can coexist with Mandarin (Guo 2004), and dialects have been considered crucial carriers of regional cultures to be protected and developed since the mid-1990s. Today, dialects are commonly observed in China and are frequently utilised in informal and private settings (Spolsky 2014) among individuals from the same speech community.

This study identified three CS patterns employed by Chinese speakers in China via Muysken's (2000, 2013) typology. Insertion was the most frequently observed pattern, followed by backflagging and alternation. No congruent lexicalisation pattern was observed (see Table 8). Insertion was the most common pattern of CS among the Mandarin-dominated speakers in mainland China in this study, accounting for more than 88% of the observed instances. Insertion only requires the use of grammar from one language, which is Mandarin in this study. In the current study, insertional CS was primarily performed from English to Mandarin, which highlighted English elements in a Mandarin matrix structure. Excerpt 1 portrays the insertion of the English word 'support' into a Mandarin-dominated sentence:

Excerpt 1 (Insertion): *English* [E]—Mandarin [M]
F1: 你们是欧洲的同事可以*support*你们的。 [M—E]
Your European colleagues can **support** you. [English translation]

Table 8 shows that word insertion was the most prominent insertion pattern, accounting for half of the total CS patterns, followed by phrase, discourse, letter, morpheme, and clause insertions. Chinese speakers in China are in the early stage of multilingualism or non-balanced multilingualism, as CS generally commences with small insertions before moving to alternation and large insertions. In this vein, congruent lexicalisation is a sign

of the most fluent multilingualism (Lipski, 2014). Backflagging, which was the second most frequently utilised CS pattern, involves the reverse insertion and adheres to the grammatical structure of the embedded language, referring to non-Mandarin varieties in the current study. Excerpt 2 shows that English sentences were backflagged in a Mandarin-dominated discourse:

Excerpt 2 (Backflagging): *English* [E]—Mandarin [M]
Both: *Hi, sister!* [E]
M: *How are you?* [E]
F: *I'm fine, thank you, and you?* [E]
M: 啊你最近皮肤真的很好诶, 状态真的非常得不错。 [M]
Ah, your skin has been really good lately. It's really good. [English translation]
F: 哦, 谢谢, 你也是! [M]
Oh, thanks, you too! [English translation]

Alternation involves the grammatical competition of two or more languages in a discourse. In Excerpt 3, Mandarin and English expressions adhered to grammatical structures. Generally, the alternation pattern indicated a low frequency of CS practices among the mainland Chinese speakers, at approximately 3%. The low frequency of alternation was predictable, as alternation may frequently occur among speakers proficient in all the languages involved with a comparable status in the community (Stell and Couto, 2012):

Excerpt 3 (Alternation): Mandarin [M]—*English* [E]
M1: 彭姐, 给你买一杯奶茶, 三分糖, *just like you, sweet enough but not too much*。 [M—E]
Sister Peng, I bought a cup of milk tea for you, light sugar, *just like you, sweet enough but not too much*. [English translation]

This result coincides with Lipski's (2014) and Halpin and Melzi's (2021) findings, where insertion is the most common CS pattern, and is frequently applied by speakers with lower proficiency in the inserted language (Stell and Couto 2012) in addition to backflagging. In contrast, alternation was scarcely employed by second-language speakers (Lipski, 2014), as Mandarin-dominated mainland Chinese speakers who have acquired dialects as a second language and English as a foreign language (Kachru 1990) prefer insertion and backflagging to alternation and congruent lexicalisation. Alternation and congruent lexicalisation require greater proficiency in all the involved languages. Alternation allows two grammars to coexist in the same discourse independently, whereas congruent lexicalisation occurs only when two languages have a similar grammar structure. Moreover, congruent lexicalisation requires languages sharing a common grammatical structure and in contact to be highly congruent in structure, either partially or entirely. Congruent lexicalisation is a frequent pattern among the most fluent multilinguals (Lipski 2014). In short, Mandarin-dominated mainland Chinese speakers are non-balanced multilinguals, which is consistent with Lipski's (2014) findings regarding English–Spanish CS practices in the American context.

The current study applied Ritchie and Bhatia's (2013) framework to elucidate the factors influencing CS practices among mainland Chinese speakers. Table 9 presents five factors examined after the content analysis of the CS practices in the 16 short videos. Another factor, convenience, was proposed in addition to the four proposed by Ritchie and Bhatia (2013). The findings demonstrated that participants' social roles and relationships were the primary factors influencing Mandarin-dominated Chinese to code-switch in China, followed by message-intrinsic considerations, sociopsychological factors, convenience, and situational factors.

As shown in Table 9, participants' social roles and relationships were the most significant factors influencing CS practices. The participants frequently engaged in CS to signify personal

Table 8 Frequencies of code-switching patterns.		
	Total number of words	**Percentage/%**
Insertion	713	88.461
Morpheme Insertion	25	3.102
Letter Insertion	51	6.327
Word Insertion	406	50.372
Phrase Insertion	125	15.509
Clause Insertion	16	1.985
Discourse Insertion	90	11.166
Alternation	27	3.350
Congruent Lexicalisation	0	0
Backflagging	66	8.189
Total CS Words	806	100

Table 9 Frequencies of factors influencing code-switching practices.

	Factors	Total words	Percentage/%
1	Social roles and relationships of participants	331	41.067
	Social roles and relationships of participants: dual/multiple identities	316	39.206
	Social roles and relationships of participants: participants	15	1.861
2	Situational factors	2	0.248
	Situational factors: discourse topic	2	0.248
3	Message-intrinsic considerations	257	31.886
	Message-intrinsic considerations: hedging	63	7.816
	Message-intrinsic considerations: idioms and deep-rooted cultural wisdom	74	9.181
	Message-intrinsic considerations: interjection	8	0.993
	Message-intrinsic considerations: paraphrasing/reiteration	23	2.854
	Message-intrinsic considerations: repetition	41	5.087
	Message-intrinsic considerations: quotation	48	5.955
4	Sociopsychological factors	165	20.471
	Sociopsychological factors: language dominance	165	20.471
5	Convenience	51	6.328
	Total words	806	100

identities or address distinct participants through purposive language mismatching or accommodation (Ritchie and Bhatia 2013). Language mismatching indicates that two speakers adhere to distinctive languages at the same comprehension level and seek to maintain identities attached to the selected languages. Speech accommodation refers to several multilingual participants attempting to use a common language for communication while realising collective identity. In Excerpt 4, M2 code-switched to signify identities through language mismatching. 'Tim' was a Mandarin-dominated Chinese male who mostly employed CS when referring to technical terms in the entire discourse. Tim suddenly changed from Mandarin to English to highlight his identity. He was an "up主", similar to a YouTuber, and possessed expertise in technology. Tim utilised an English name instead of his real Mandarin name to conceal his original identity while demonstrating an international identity:

Excerpt 4 (Participants' Social Roles and Relationships: Dual or Multiple Identities):
Mandarin [M]—*English* [*E*]
M2: 各位好, 我是*Tim*。 [M—*E*]
Hi, everyone, I'm *Tim*. [English translation]

Participant-related factors, including age, gender, religion, social class, and background, significantly impacted code choices (Ritchie and Bhatia 2013; Ndimande-Hlongwa and Ndebele 2014; Sardar et al. 2015; Alkhawaldeh 2019). Excerpt 5 shows that F2 was Mandarin-dominated when addressing F1, but switched to English when directing global colleagues. Specifically, the addressee influenced the speaker to code-switch:

Excerpt 5 (Participants' Social Roles and Relationships: Participants):
Mandarin [M]—*English* [*E*]
F1: 林风现在正在开他们全球的一个销售的会议。 [M]
Lin Feng is in one of their global sales meetings right now. [English translation]
F2: *Ha ha, morning Kris, hi, everyone. Besides, I was also a bit occupied by the existing client*. [*E*]

Certain languages were considered more suitable for certain topics or settings. Discourse topics, which are Ritchie and Bhatia's (2013) situational factors, influenced Mandarin-dominated mainland Chinese speakers to code-switch in daily interactions. Certain specific words constantly appeared regarding certain topics despite Mandarin equivalences. Excerpt 6 shows a term, 'CT', that refers to the Mandarin expression '计算机断层扫描 (computed tomography)'. 'CT' is more commonly utilised by

Chinese people in daily life than '计算机断层扫描 (computed tomography)'. The participants would intuitively and continuously code-switch when discussing certain topics or situations (Stapa and Khan 2016; Ai Rousan and Merghmi 2019):

Excerpt 6 (Situational Factors: Discourse Topic): Mandarin [M]—*English* [*E*]
M1: 包括做了这次肺部的低剂量的螺旋*CT*它就会扫到你这个上半身的每一个
脏器, 过后医生就跟我说, 可能他们在*CT*的结果里面会看到我肺部会有
一些结节。 [M—*E*]
I did this low-dose spiral *CT* scan of the lung which would scan every organ in your upper body, and then the doctor told me that maybe they would see some nodules in my lungs on the *CT*. [English translation]

Hedging, idioms, deep-rooted cultural wisdom, interjection, paraphrasing, repetition, and quotation were message-intrinsic considerations influencing Mandarin-dominated Chinese to code-switch.

Hedging is defined as taboo suppression, deintensification, or vague expressions (Ritchie and Bhatia 2013). Excerpt 7 illustrates an example of hedging, wherein the speaker reminisced about their first day as a salesperson. The speaker employed English expressions, including 'balance' and 'low', instead of the Mandarin counterparts '平衡 (balance)' and '低级 (low)'. The hesitation markers before 'balance' and 'low', such as '什么 (any)' and '啧 (tut-tut)', indicated that the speaker was thinking and making the language selections deliberately. 'Balance' and 'low' were code-switched to decrease the degree of seriousness, as directly employing '平衡 (balance)' and '低级 (low)' would convey a more solemn tone. In this context, CS served as hedging:

Excerpt 7 (Message-intrinsic Considerations: Hedging): Mandarin [M]—*English* [*E*]
F2: 嗯。好多人说, 哎呀女销售或者是销售你要*balance*好你的生活, 其实我觉
得挺扯淡的。我觉得没有什么*balance*。(…) 就是我一直觉得我, 销售有点
难为情, 好像不是特别好, 反正就是挺, 啧, *low*的。 [M—*E*]
Emm. A lot of people say, oh saleswomen or salesmen have to *balance* their lives, I think it's pretty bullshit. I don't think there is any *balance*. (…) I always feel embarrassed about my job, always think it's not so good, anyway, I feel it's too, tut-tut, *low*. [English translation]

Excerpt 8 is an example of code-switching highlighting the influence of idioms and deep-rooted cultural wisdom. M applied

a fragrance to his lips and felt uncomfortable with the smell after several seconds. F laughed at M, uttering the Chinglish idiom 'no zuo no die' (this would not happen if you were not so dramatic), which was formed by referring to the Chinese idiom structure 'ABAC'. In particular, 'zuo' is the Mandarin pronunciation of the word '作 (dramatic)' and 'die' is the English translation of the Mandarin word '死 (die)'. The combination of negative English words, Mandarin pronunciation, and English translation formed the idiom 'no zuo no die':

Excerpt 8 (Message-intrinsic Considerations: Idioms and Deep-Rooted Cultural Wisdom): Mandarin [M]—*Chinglish* [*Ch*]

M: 香精味道。 [M]
The smell of fragrance. [English translation]
F: 这个还是不能涂嘴上的。 [M]
This can't be put on lips. [English translation]
M: 对, 就不能涂嘴上。哎哟我的妈呀。 [M]
Yes, it can't. Oh, my God! [English translation]
F: *No zuo no die.* [*Ch*]
This wouldn't happen if you weren't so dramatic. [English translation]

Interjection or sentence filler is another CS function. Multilinguals frequently mix several modal particles from another language. The word 'wow' in Excerpt 9 mixed an English interjection into a Mandarin-dominated sentence. The pronunciation of 'wow [waʊ]' was audible in the video. Only a few Mandarin interjections are available, including 哇 [wà], 哇哦 [wà'o], or 喔 [wò], to express astonishment. All of these expressions have pronunciations and intonations that differ from the English interjection 'wow [waʊ]':

Excerpt 9 (Message-intrinsic Considerations: Interjection): Mandarin [M]—*English* [*E*]

M: *Wow*, 这么高端? 快给我看看你买了什么。 [*E*—M]
Wow, so high-end? Show me what you got. [English translation]

Another CS function under message-intrinsic considerations is paraphrasing. In Excerpt 10, two speakers, M1 and M2, discussed their respective universities and majors. M1 stated, '那我呢是毕业于中国药科大学的药学英语' (I graduated from China Pharmaceutical University with a degree in pharmaceutical English) in a fully Mandarin sentence, whereas M2 laughed and code-switched to an English explanation 'to learn English' (要学英语). M2 intended to paraphrase and render the mysterious and strange expression '药学英语 (pharmaceutical English)' friendly to the audience. M2 not only paraphrased but also created certain humour effects, as the pronunciation of '药学英语 (yào xué yīng yǔ, pharmaceutical English)' resembled that of '要学英语 (yào xué yīng yǔ, to learn English)'. M2's code-switched utterances achieved both paraphrasing and humour functions:

Excerpt 10 (Message-intrinsic Considerations: Paraphrasing): Mandarin [M]—*English* [*E*]

M1: 那我呢是毕业于中国药科大学的药学英语。药学英语, 大家听起来是不是很
神秘又陌生? [M]
I graduated from China Pharmaceutical University with a degree in pharmaceutical English. Pharmaceutical English, does it sound mysterious and strange to you? [English translation]
M2: 呵呵呵呵。 *To learn English*. [M—*E*]
Hehehehe. *To learn English*. [English translation]

Repetition, which serves as a CS function, is easily recognised among discourses. In Excerpt 11, two speakers discussed the meaning of 'to C'. F1 asked F2 to explain it, and F2 responded that the meaning differed from that of the current job: sales to enterprise-level customers. According to F1, 'to C' was an abbreviation of 'to customer', which refers to sales as targeting an individual rather than a team. The second 'to C' was a repetition of the first 'to C':

Excerpt 11 (Message-intrinsic Considerations: Repetition): Mandarin [M]—*English* [*E*]

F1: *To C* 的那个销售叫什么? [*E*—M]
What does sales called *to C* mean? [English translation]
F2: 其实都叫销售。比如说柜姐, 也是销售。那企业级客户的意思就是… [M]
They are all called sales. For example, the counter saleswoman is also called sales. Then enterprise-level customer means… [English translation]
F1: 就是说*to C*的销售它是说服一个人, 你们是说服一个团队。 [M—*E*]
That means *to C* is to convince one person. You are to convince a team. [English translation]

Direct quotation or reported speech triggered speakers to switch across languages. The term 'fuck-you money' in Excerpt 12 was directly quoted from an interview with Lucy Liu, wherein Liu used the term to refer to savings that serve as a cushion when a person quits a job in front of their supervisor. The corresponding connotations were also present when participants quoted other individuals:

Excerpt 12 (Message-intrinsic Considerations: Direct Quotation):
Mandarin [M]—*English* [*E*]

M1: 我觉得就有点像刘玉玲说的, 呃, 因为工作了这么多年嘛, 我们算是手上有
一点点*fuck-you money*的。 [M—*E*]
I think it's like Lucy Liu said, well, after working for so many years, we have a little *fuck-you money* on our hands. [English translation]

Language dominance determined the quality and quantity of CS. Certain expressions were effortlessly and directly switched into other languages without a translated form. 'Supreme', referenced in Excerpt 13, is the English name of a sportswear brand without a Mandarin translation. English is the dominant language of the brand name, and the participants were required to code-switch the brand name in the Mandarin-dominated discourse:

Excerpt 13 (Sociopsychological Factors: Language Dominance):
Mandarin [M]—*English* [*E*]

F: *Supreme*的看到了吗? [M—*E*]
It's *Supreme*. See? [English translation]
M: 啊*Supreme*的。这个是塑料的碗吗? [M—*E*]
Ah, *Supreme*. Is this a plastic bowl? [English translation]

Convenience is a common rationale for CS, as code-switched language is shorter or simpler compared to the factor of language dominance without Mandarin equivalence (Leung and Chan 2016). Several inserted abbreviations in the Chinese context were possibly motivated by convenience. Excerpt 14 shows how 'LED' was more convenient than the Mandarin equivalent of '发光二极管':

Excerpt 14 (Convenience): Mandarin [M]—*English* [*E*]

M2: 它亮度非常非常高, 属于高功率的*LED*灯。 [M—*E*]
It is very bright and belongs to a high-power *LED* lamp. [English translation]

Five factors triggering CS were examined among Mandarin-dominated Chinese speakers in China. The participants' social roles and relationships related to personal identities and addressees were the key elements contributing to CS practices. These findings paralleled those of Choi (2014), Priror and Masruddin (2019) and Gross et al. (2022). The secondary vital factor was linguistic factors of hedging, idioms and deep-rooted cultural wisdom, interjections, paraphrasing, repetition, and quotation, followed by sociopsychological factors, convenience,

and situational factors. The results confirmed that Ritchie and Bhatia's (2013) framework could adequately explain CS practices among Chinese speakers in China. Another new factor proposed by Leung and Chan (2016)—convenience—was discovered to expand Ritchie and Bhatia's (2013) framework to elucidate the CS practices in China.

Conclusion

This study provides insights into the CS behaviours of Mandarin speakers in China, revealing a predominance of insertional patterns. The findings contribute to understanding multilingual practices in China and set the stage for further research into the complexities of language use in multilingual societies.

It revealed that English was the most frequently code-switched language in addition to Mandarin in the Mandarin-dominated discourses among mainland Chinese speakers, followed by various regional dialects, such as the Beijing, Northeastern and Sichuan dialects, spoken equivalently by distinct dialectal communities. Based on Muysken's (2000, 2013) typology, only three of four patterns (insertion, alternation, and backflagging) were identified among the Mandarin-dominated mainland Chinese speakers. Insertion was the most prominent pattern, accounting for more than 88%. The insertional pattern can be further divided into smaller linguistic units. Word insertion was the largest portion of the total CS patterns, followed by phrase, discourse, letter, morpheme, and clause insertions. Backflagging was the second most common pattern, accounting for about 8% of CS practices. Alternation was applied at the least. No congruent lexicalisation pattern has been discovered.

The dominant insertional CS pattern suggests that the multilingual capabilities within China's Mandarin-speaking community are in a formative phase characterised by a reliance on single-language grammar structures. The high frequencies of insertion and backflagging suggest that the Mandarin-dominated mainland Chinese speakers were less proficient in foreign languages or dialects. Insertion and backflagging require only limited ability in one language, whereas alternation and congruent lexicalisation require greater proficiency in grammatical structures in all the languages involved. Moreover, the Chinese speakers were not proficient in all the insertion types but only in word insertion, which is as same as Halpin and Melzi's (2021) findings among preschoolers. This finding indicates a low multilingualism level, as CS generally commences with small insertions before moving to large insertions and alternation, with congruent lexicalisation as the sign of the most fluent multilingualism (Lipski 2014).

The current study also proved the feasibility of Ritchie and Bhatia's (2013) framework and supported an emergent factor proposed in Leung and Chan's (2016) empirical studies. Five CS factors were examined, including four factors from Ritchie and Bhatia's (2013) framework and another factor supplementing the framework. The most common factor was participants' social roles and relationships, followed by message-intrinsic considerations, sociopsychological factors, convenience, and situational factors. The results suggested that the CS practices of the Mandarin-dominated mainland Chinese speakers were significantly influenced by personal factors (Choi 2014; Prirol and Masruddin 2019; Gross et al. 2022). The dominance of personal factors also implies that China is in the primary stage of multilingualism, as those factors are closely related to the individual per se rather than situational and linguistic factors.

This study has several limitations. It is limited by its focus on Mandarin-dominated speakers and its relatively small sample size. It only partially reflects the language practices of Mandarin-dominated mainland Chinese speakers, and may not provide a comprehensive depiction of CS practices in diverse circumstances. Secondly, utilising the current methodology may only partially reveal the language practices of mainland Chinese speakers. Although this method stands out for obtaining online short video data more quickly and easily, participants are more willing and find it more convenient to select videos rather than record their conversations, especially during the pandemic. Thirdly, the small sample size may be insufficient for a thorough understanding of CS patterns of alternation and congruent lexicalisation. Other factors, such as macro-level sociocultural interactions and micro-level psychological attitudes, resulting in the primary stage of multilingualism in China, were not investigated.

Based on the findings and limitations listed above, future studies could explore CS practices among non-Mandarin dominant speakers and incorporate larger, more diverse samples to enhance the generalisability of the findings. This study also raised awareness of the language practices among non-balanced multilingual communities. Besides, further explorations are encouraged on language policy, education and social interactions among the communities. More potential factors, such as language purism, sociocultural factors, personal language proficiency, and psychological attitude, are to be examined.

References

Adamson B, Feng A (eds) (2021) Multilingual China: National, minority and foreign languages, 1st edn. Routledge, London and New York

Ai Rousan R, Merghmi K (2019) A conversational analysis of Arabic-French code switching in Algerian TV talk shows. Jordan J Mod Lang Lit 11:247–271

Al-Daher ZY (2021) Necode-switching to English amongst Arabic-speaking Jordanians in Canada. Jordan J Mod Lang Lit 13:477–499. https://doi.org/10.47012/jjmll.13.3.7

Albirini A, Chakrani B (2017) Switching codes and registers: An analysis of heritage Arabic speakers' sociolinguistic competence. Int J Biling 21:317–339. https://doi.org/10.1177/1367006915626587

Alkhawaldeh A (2019) Code-switching between Arabic and English: Reasons, types and attitudes as expressed by EFL female students at Imam Muhammad Ibn Saud Islamic University. Int J English Linguist 9:135. https://doi.org/10.5539/ijel.v9n6p135

Bidnoshyia Y, Dyka L (2022) Dialects in the current sociolinguistic situation in Ukraine. Cogn Stud 2022. https://doi.org/10.11649/cs.2652

Carstens S, Ang LH (2019) Conversational code switching: Languaging Chinese identities in multilingual Malaysia. Asian J Soc Sci 47:508–533. https://doi.org/10.1163/15685314-04704005

Chan KLR (2018) Being a 'purist' in trilingual Hong Kong: Code-switching among Cantonese, English and Putonghua. Linguist Res 35:75–95. https://doi.org/10.17250/khisli.35.1.201803.003

Choi J (2014) Code-switching in an institutional setting: Negotiating social roles in bilingual encounters in Guatemala. Circ Linguist Apl a la Comun 60:36–61. https://doi.org/10.5209/rev_CLAC.2014.v60.47443

Demirçay D, Backus A (2014) Bilingual constructions: Reassessing the typology of code-switching. Dutch J Appl Linguist 3:30–44. https://doi.org/10.1075/dujal.3.1.02dem

Dewaele JM, Zeckel I (2016) The psychological and linguistic profiles of self-reported code-switchers. Int J Biling 20:594–610. https://doi.org/10.1177/1367006915575411

Eberhard DM, Simons GF, Fennig CD (eds) (2021) Ethnologue: Languages of the world, 24th edn. Dallas, Texas: SIL International

Fakeye DO (2012) Motivational factors for code alternation in pre-service teachers' verbal communication in Oyo and Ekiti States. Asian Soc Sci 8:149–154. https://doi.org/10.5539/ass.v8n8p149

Fishman J (1965) Who speaks what language to whom and when? La Linguist 1:67–88

Francis N (2016) Language and dialect in China. Chinese Lang Discourse 7:136–149. https://doi.org/10.1075/cld.7.1.05fra

Gross MC, López González AC, Girardin MG, Almeida AM (2022) Code-switching by Spanish–English bilingual children in a code-switching conversation sample: Roles of language proficiency, interlocutor behavior, and parent-reported code-switching experience. Languages 7. https://doi.org/10.3390/languages7040246

Gumperz JJ (1982) Discourse strategies, 1st edn. Cambridge University Press, Cambridge

Guo L (2004) The relationship between Putonghua and Chinese dialects. In: Zhou M (ed) Language policy in the People's Republic of China: Theory and practice since 1949. Kluwer Academic Publishers, Amsterdam, pp 45–53

Halpin E, Melzi G (2021) Code-switching in the narratives of dual-language Latino preschoolers. Int J Biling Educ Biling 24:1271–1287. https://doi.org/10.1080/13670050.2018.1553928

Hoffmann C (2014) Introduction to bilingualism. Routledge, London and New York

Holmes J (ed) (2013) An introduction to sociolinguistics, 4th edn. Routledge, London and New York

Kachru BB (1990) World Englishes and applied linguistics. World Englishes 9:3–20

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159. https://doi.org/10.2307/2529310

Leung CH, Chan WTY (2016) Sociolinguistic phenomenon of code mixing in Hong Kong: From a perspective of marketing communications. Humanit Soc Sci Rev 4:20–26. https://doi.org/10.18510/hssr.2016.413

Li DCS (2006) Chinese as a lingua franca in greater China. Annu Rev Appl Linguist 26:149–176. https://doi.org/10.1017/S0267190506000080

Li DCS, Elly CY (2002) One day in the life of a "purist". Int J Biling 6:147–203. https://doi.org/10.1177/13670069020060020301

Li W (2011) Multilinguality, multimodality, and multicompetence: Code- and mode switching by minority ethnic children in complementary schools. Mod Lang J 95:370–384. https://doi.org/10.1111/j.1540-4781.2011.01209.x

Lipski JM (2014) Spanish-English code-switching among low-fluency bilinguals: Towards an expanded typology. Socioling Stud 8:23–55. https://doi.org/10.1558/sols.v8i1.23

Maďarová S (2018) Use and perception of code switching by students and instructors in the second language classroom in the US. Onomazein 150–177. https://doi.org/10.7764/ONOMAZEIN.48.07

Maher JC (2017) Multilingualism: A very short introduction, 1st edn. Oxford University Press, New York

Malik L (ed) (1994) Sociolinguistics: A study of code-switching. Anmol Publications PVT. LTD., New Delhi

May LM, Aziz AA (2020) Teachers' use of code-switching in ESL classrooms at a Chinese vernacular primary school. Int J English Lang Lit Stud 9:41–55. https://doi.org/10.18488/journal.23.2020.91.41.55

Muysken P (2000) Bilingual speech: A typology of code-mixing. Cambridge University Press, Cambridge

Muysken P (2013) Language contact outcomes as the result of bilingual optimization strategies. Bilingualism 16:709–730. https://doi.org/10.1017/S1366728912000727

Myers-Scotton C (1993) Duelling languages: Grammatical structure in code-switching. Clarendon Press, Oxford

Myers-Scotton C (2002) Contact linguistics: Bilingual encounters and grammatical outcomes. Oxford University Press, New York

Ndimande-Hlongwa N, Ndebele H (2014) Digging deep into IsiZulu-English code-switching in a peri-urban context. Lang Matters Stud Lang Africa 45:237–256. https://doi.org/10.1080/10228195.2014.910248

Poplack S (1980) Sometimes I start a sentence in Spanish y termino en español: Toward a typology of code-switching. Linguistics 18:581–618

Prirol A, Masruddin (2019) Reasons and motivation of Islamic scholar for using code-switching as strategy in delivering a speech (Da'wah). Asian EFL J 25:265–275

Ritchie WC, Bhatia TK (2013) Social and psychological factors in language mixing. In: Bhatia TK, Ritchie WC (eds) The handbook of bilingualism and multilingualism, 2nd edn. Blackwell Publishing, Ltd, pp 375–390

Sardar SS, Mahdi AAA, Mohd YMS (2015) Code-switching in daily conversations among Iraqi students in Malaysia. Arab World English J 6:309–319. https://doi.org/10.24093/awej/vol6no3.20

Smolak E, de Anda S, Enriquez B, Poulin-Dubois D, Friend M (2020) Code-switching in young bilingual toddlers: A longitudinal, cross-language investigation. Bilingualism 23:500–518. https://doi.org/10.1017/S1366728919000257

Spolsky B (2014) Language management in the People's Republic of China. Language (Baltim) 90:e165–e179. https://doi.org/10.1353/lan.2014.0075

Stapa SH, Khan NNBS (2016) Functions of code-switching: A case study of a mixed Malay-Chinese family in the home domain. Pertanika J Soc Sci Humanit 24:181–194

Stell G (2010) Ethnicity in linguistic variation: White and Coloured identities in Afrikaans-English code-switching. Pragmatics 20:425–447. https://doi.org/10.1075/prag.20.3.06ste

Stell G, Couto CP (2012) Code-switching practices in Luxembourg's Portuguese-speaking minority: A pilot study on the distinctive characteristics of an immigrant community's code-switching practices within a trilingual majority. Zeitschrift fur Sprachwiss 31:153–185. https://doi.org/10.1515/zfs-2012-0004

Tramutoli L (2021) The code-mixing of the Senegalese migrants in Italy. Int J Biling 25:1235–1262. https://doi.org/10.1177/13670069211008243

Vaughan J (2021) Enduring and contemporary code-switching practices in northern Australia. Languages 6:1–25. https://doi.org/10.3390/languages6020090

Wardhaugh R, Fuller JM (eds) (2015) An introduction to sociolinguistics, 7th edn. John Wiley & Sons, Inc., Chichester

Wu D, Cai L, Liang L, Li H (2022) Patterns and predictors of code-switching in Singapore preschoolers: A corpus-based study. Int J Biling Educ Biling 25:2933–2948. https://doi.org/10.1080/13670050.2021.1997903

Yahiaoui R, Aldous MJ, Fattah A (2021) Functional and sociocultural attitudes of code-switching and its relation to the meaning-making process: The case of dubbing Kim Possible into Arabic. Int J Biling 25:1349–1368. https://doi.org/10.1177/13670069211019482

Yu M (2021) 中国语言文字概况(2021年版) [Overview of Chinese language and characters (2021 edition)]. In: Gov. portal Minist. Educ. People's Repub. China. http://www.moe.gov.cn/jyb_sjzl/wenzi/202108/t20210827_554992.html. Accessed 19 Feb 2022

Zhang H, Cai S (2020) Putonghua vs. minority languages: Distribution of language laws, regulations, and documents in mainland China. Ethn Racial Stud 44:2574–2594. https://doi.org/10.1080/01419870.2020.1828598

Zhang W (2005) Code-choice in bidialectal interaction: The choice between Putonghua and Cantonese in a radio phone-in program in Shenzhen. J Pragmat 37:355–374. https://doi.org/10.1016/j.pragma.2004.10.007

Zhang W, Bolton K, Botha W (2020) English in the People's Republic of China. In: Nelson CL, Proshina ZG, Davis DR (eds) The handbook of World Englishes, 2nd edn. John Wiley & Sons, Inc., New Delhi

Zhong X, Ang LH, Sharmini S (2023) Systematic literature review of conversational code-switching in multilingual society from a sociolinguistic perspective. Theory Pract Lang Stud 13:318–330. https://doi.org/10.17507/tpls.1302.06

Author contributions

All the authors have read and approved the final manuscript and agree to be accountable for all aspects of the research.

Competing interests

The authors declare no competing interests.

Ethical approval

The researchers requested ethical approval from the Ethics Committee for Research Involving Human Subjects at Universiti Putra Malaysia. Universiti Putra Malaysia granted ethical approval to the research (Ref No: JKEUPM-2023-382). Therefore, the research protocol and procedures of the current study were screened and approved by the Ethics Committee for Research Involving Human Subjects at Universiti Putra Malaysia. The approval from the Ethics Committee for Research Involving Human Subjects ensures that the study's data collection methods and ethical considerations meet the essential standards for conducting research involving human participants.

Informed consent

The informed written consent form was obtained from the participant before the data collection.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-03534-z.

Correspondence and requests for materials should be addressed to Lay Hoon Ang.

Reprints and permission information is available at http://www.nature.com/reprints

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